PCT WORLD INTELLECTUAL PROPERTY ORGA. ATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07H 1/00, 1/02, C07C 255/11, 255/49 — A1

(11) International Publication Number: WO 00/18778

(43) International Publication Date: 6 April 2000 (06.04.00)

(21) International Application Number: PCT/US99/22436

(22) International Filing Date: 28 September 1999 (28.09.99)

(30) Priority Data:
60/102,299 29 September 1998 (29.09.98) US

(71) Applicant: PHYLOS, INC. [US/US]; 128 Spring Street, Lexington, MA 02421 (US).

(72) Inventors: LOHSE, Peter; 50 Golden Ball Road, Weston, MA 02493 (US). KUIMELIS, Robert, G.; 21 Malbert Road, Brighton, MA 02135 (US).

(74) Agent: ELBING, Karen, L.; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110–2214 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SYNTHESIS OF CODON RANDOMIZED NUCLEIC ACIDS

(57) Abstract

A method for generating a selected set of codons is disclosed; the method includes the steps of: (a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or mixture thereof, where a subset A of the first set is protected with a protecting group A', and a subset B of the first set is protected with a protecting group B', where A' and B' are orthogonal protecting groups; (b) selectively removing the protecting group A' from subset A; (c) coupling the products of step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the second set is protected with protecting group A'; (d) optionally removing protecting group A' from the products of step (c); (e) optionally coupling the products of step (d) with a third set of mononucleosides, where the third set is protected with protecting group A'; (f) selectively removing the protecting group B' from subset B; (g) coupling the products of step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the fourth set is protected with protecting group A' or protecting group B'; (h) optionally selectively removing protecting group B' from the products of step (g); and (i) optionally coupling the products of step (h) with a fifth set of mononucleosides, to yield a selected set of codons.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

SYNTHESIS OF CODON RANDOMIZED NUCLEIC ACIDS

Background of the Invention

The invention relates to methods for chemically synthesizing DNA or RNA.

Pharmaceutical research relies, in part, on the identification of novel proteins with desired functions and properties. To identify proteins or peptides with improved properties, derivatives of known proteins and peptides can be prepared using methods such as oligonucleotide-directed mutagenesis. Proteins with desired functions can also be selected from pools of randomly synthesized proteins, including proteins which are generated from random DNA template libraries.

DNA libraries, in turn, may also be generated using a variety of techniques. Such DNA libraries can be synthesized on a solid support (e.g., a CPG support), in a liquid phase, or in a combination solid-liquid phase (e.g., a PEG support). Most commonly, DNA libraries are prepared using a standard DNA synthesizer and a random mixture of all 4 nucleotides in each coupling step. By this approach, the trinucleotides, or codons, that correspond to the different amino acids, are randomly generated. This codon randomized DNA can then be transcribed into RNA, which is in turn used to synthesize polypeptides; the approach described above thus provides a means for generating a wide variety of DNA sequences and proteins products.

Although it is commonly utilized, the random generation of DNA by conventional techniques can have disadvantages. For example, methods that rely on completely random generation of codons generally suffer from limited control over the synthesis of polypeptides generated from this DNA. The presence of weakly expressed codons in the random product mixture, for

example, lowers the efficiency with which the DNA is translated. Furthermore, a small subset of randomly generated codons (approximately 3 out of 64) corresponds to a stop codon. As the presence of stop codons terminates protein synthesis, protein libraries generated from randomly generated DNA templates can sometimes exhibit low yields of full-length proteins. In addition, methods that rely on the completely random generation of DNA do not allow for a bias for a selected group of amino acids, for example, hydrophobic amino acids.

Summary of the Invention

In one aspect, the invention features a method for generating a selected set of codons; the method includes the steps of: (a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where a subset A of the first set is protected with a protecting group A', and a subset B of the first set is protected with a protecting group B', where A' and B' are orthogonal protecting groups; (b) selectively removing protecting group A' from subset A; (c) coupling the products of step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the second set is protected with protecting group A'; (d) optionally removing protecting group A' from the products of step (c); (e) optionally coupling the products of step (d) with a third set of mononucleosides, where the third set is protected with protecting group A'; (f) selectively removing protecting group B' from subset B; (g) coupling the products of step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the fourth set is protected with protecting group A' or protecting group B'; (h) optionally selectively removing protecting group B' from the products of step (g); and (i) optionally coupling the products of step (h) with a fifth set of mononucleosides, to yield a selected set of codons.

In preferred methods, the selected set includes at least one codon

corresponding to each of the 20 naturally-occurring amino acids; preferably, each of these codons corresponds to a highly expressed codon for one of the naturally-occurring amino acids. The selected set may also consist of trinucleotides coding only for a class of amino acids, e.g., hydrophobic amino acids, hydrophilic amino acids, basic amino acids, or acidic amino acids. In another preferred method, the selected set may consist of trinucleotides coding for a mixture of amino acids, e.g., acidic and basic amino acids.

Preferably, fewer than 3% of the codons correspond to a stop codon; more preferably, fewer than 2%, 1%, 0.5% or 0.1%, of the codons correspond to a stop codon. In preferred methods, steps (a) to (i) take place in the same reaction vessel; in addition, protecting groups A' and B' are two different groups and are preferably chosen from an acid-cleavable protecting group (for example a dimethoxytrityl group), a base-cleavable protecting group (for example, a fluorenylmethyloxycarbonyl group), or a fluoride-cleavable protecting group (for example, a silyl group). In other preferred methods, each of the codons terminates in a cytidine or a guanosine residue.

In a second aspect, the invention features a method for generating an oligonucleotide from a selected set of codons; the method includes the steps of: (a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where a subset A of the first set is protected with a protecting group A', and a subset B of the first set is protected with a protecting group B', where A' and B' are orthogonal protecting groups; (b) selectively removing protecting group A' from subset A; (c) coupling the products of step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the second set is protected with protecting group A'; (d) optionally removing protecting group A' from the products of step (c); (e) optionally coupling the products of step (d) with a third set of mononucleosides, where the third set is protected with protecting group A'; (f)

selectively removing protecting group B' from subset B; (g) coupling the products of step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the fourth set is protected with protecting group A' or protecting group B'; (h) optionally selectively removing protecting group B' from the products of step (g); (i) optionally coupling the products of step (h) with a fifth set of mononucleosides; (j) removing the protecting groups from the products of step (g) or (i); and (k) repeating steps (a) to (j) until an oligonucleotide with the desired length is achieved. Preferably, steps (a) to (k) take place in the same reaction vessel.

In a third aspect, the invention features a method for generating a selected set of codons including the steps of: (a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where a subset A of the first set is protected with a protecting group A', a subset B of the first set is protected with a protecting group B', and a subset C of the first set is protected with a protecting group C', where A', B', and C' are orthogonal protecting groups; (b) selectively removing the protecting group A' from the subset A;(c) coupling the product formed in step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the second set is protected with the protecting group A'; (d) optionally removing the protecting group A' from the products of step (c); (e) optionally coupling the products of step (d) with a third set of mononucleosides, where the third set of mononucleosides is protected with the protecting group A'; (f) selectively removing the protecting group B' from the subset B; (g) coupling the products formed in step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the fourth set is protected with the protecting group A' or the protecting group B'; (h) optionally selectively removing the protecting group B' from the products of step (g); (i) optionally coupling the products of step (h) with a fifth set of

mononucleosides, where the fifth set is protected with protecting group A'; (j) selectively removing the protecting group C' from the subset C; (k) coupling the products formed in step (j) with a sixth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where a subset of the sixth set is protected with the protecting group C', and the remainder of the sixth set is protected with protecting group B'; (l) optionally selectively removing the protecting group B' from the products of step (k); (m) optionally coupling the products of step (l) with a seventh set of mononucleosides, where the seventh set is protected with protecting group A' or protecting group B'; (n) selectively removing the protecting group C' from the products of step (m); and (o) coupling the products of step (n) with an eighth set of mononucleosides, to yield a selected set of codons.

In preferred methods, steps (a) to (o) take place in the same reaction vessel. In addition, one of the protecting groups A', B', and C' is preferably an acid-cleavable protecting group (for example a dimethoxytrityl group), another of the protecting groups A', B', and C' is preferably a base-cleavable protecting group (for example, a fluorenylmethyloxycarbonyl group), and the last of the protecting groups A', B', and C' is preferably a fluoride-cleavable protecting group (for example, a silyl group).

In a fourth aspect, the invention features a method for generating an oligonucleotide from a selected set of codons including the steps of: (a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where a subset A of the first set is protected with a protecting group A', a subset B of the first set is protected with a protecting group B', and a subset C of the first set is protected with a protecting group C', where A', B', and C' are orthogonal protecting groups; (b) selectively removing the protecting group A' from the subset A; (c) coupling the product formed in step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture

thereof, where the second set is protected with the protecting group A'; (d) optionally removing the protecting group A' from the products of step (c); (e) optionally coupling the products of step (d) with a third set of mononucleosides, where the third set is protected with the protecting group A'; (f) selectively removing the protecting group B' from the subset B; (g) coupling the products formed in step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the fourth set is protected with the protecting group A' or the protecting group B'; (h) optionally selectively removing the protecting group B' from the products of step (g); (i) optionally coupling the products of step (h) with a fifth set of mononucleosides, where the fifth set is protected with protecting group A'; (j) selectively removing the protecting group C' from the subset C; (k) coupling the products formed in step (j) with a sixth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where a subset of the sixth set is protected with the protecting group C', and the remainder of the sixth set is protected with protecting group B'; (l) optionally selectively removing the protecting group B' from the products of step (k); (m) optionally coupling the products of step (l) with a seventh set of mononucleosides, where the seventh set is protected with protecting group A' or protecting group B'; (n) selectively removing the protecting group C' from the products of step (m); (o) coupling the products of step (n) with an eighth set of mononucleosides; (p) removing the protecting groups from the products of step (o); and (q) repeating steps (a) to (p) until an oligonucleotide with the desired length is achieved. Preferably, steps (a) to (q) take place in the same reaction vessel.

In a fifth aspect, the invention features a method for generating, in the same reaction vessel, a selected set of codons; the method includes the steps of: (a) providing a first set of mononucleosides, mononucleotides, or dinucleotides, or mixture thereof; (b) adding a second set of mononucleosides,

mononucleotides, dinucleotides, or a mixture thereof; (c) optionally adding a third set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof; and (d) optionally repeating step (c) to yield a selected set of codons. The selected set includes at least one codon having A or G at the third codon position; fewer than 3% of the codons in the selected set correspond to a stop codon.

In preferred methods, the selected set includes at least one codon for each of the 20 naturally-occurring amino acids; preferably, each codon corresponds to a highly-expressed codon for one of the naturally-occurring amino acids. In other preferred methods, the selected set may consist of one class of codons, e.g., hydrophobic amino acids. In another preferred method, the selected set may consist of trinucleotides coding for a mixture of amino acids, e.g., acidic and basic amino acids. Preferably, fewer than 2% of the codons correspond to a stop codon; more preferably, fewer than 1%, 0.5%, or 0.1%, of the codons correspond to a stop codon. In still other preferred methods, each of the codons terminates in a cytidine or a guanosine residue.

Any combination of couplings of mononucleosides, mononucleotides, and dinucleotides may be used to generated codons, which are trinucleotides. For example, dinucleotides may be coupled with mononucleosides. Dinucleotides would not be coupled with dinucleotides, as that would generated tetranucleotides.

In a sixth aspect, the invention features a method for generating an oligonucleotide from a selected set of codons. The method includes the steps of: (a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof; (b) adding a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof; (c) optionally adding a third set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof; and (d) optionally repeating step (c) to yield

a selected set of codons that includes at least one codon having A or G at the third codon position and in which fewer than 3% of the codons correspond to a stop codon. Steps (a), (b), (c), and (d) occur in the same reaction vessel; these steps are repeated until an oligonucleotide of the desired length is achieved. Preferably, the selected set includes at least one codon for each of the 20 naturally-occurring amino acids, and fewer than 2% of the codons correspond to a stop codon.

By "nucleoside" is meant any sugar-base moiety, including sugar-base moieties in which one or more nitrogen atoms of the nitrogenous bases are protected, and/or in which the 5'-OH of the sugar is protected. "Nucleosides" also include nucleoside phosphoramidites and protected phosphoramidites.

By "nucleotide" is meant any sugar-phosphate-base moiety, as well as any derivatized sugar-phosphate-base moiety. One or more nitrogen atoms of the nitrogenous bases can be protected, and/or the 5'-OH of the sugar can be protected. Dinucleotides can include dinucleotide phosphoramidites; in addition, the internucleotide linkage may be protected.

By "oligonucleotide" is meant either a DNA sequence or an RNA sequence; by "nucleic acid" is meant either DNA or RNA.

By "highly-expressed codons" are meant the codons present in higher than normal abundance in highly expressed genes.

By "stop codon" is meant one of the DNA codons TAA, TGA, and TAG; and the RNA codons UAA, UGA, and UAG.

By a "selected set of codons" is meant a set of trinucleotide sequences where each trinucleotide has an assigned representation in the set. For example, a selected set of codons may be a set that contains at least one codon for each of the naturally occuring amino acids (e.g., AAC : CAC : GAC : TAC: ACC : CCC : GCC : TCC : AGC : CGC : GGC : TGC : ATC : CTC : GTC : TTC : AAG : CAG : GAG : TGG : ATG =

1:1:1:1:1:1:1:1:1:1:1:1:1:1:1:1:1:1:1:1:1). Alternatively, a selected set of codons may be a set that contains at least one codon for each of the naturally occurring amino acids, and in which some hydrophobic amino acids (e.g., Val, Leu, Ile, Phe) are twice as abundant (e.g., AAC : CAC : GAC : TAC : ACC : CCC : GCC : TCC : AGC : CGC : GGC :TGC : ATC : CTC : GTC : TTC : AAG : CAG : GAG : TGG : ATG = 1:1:1:1:1:1:1:1:1:1:2:2:2:2:2:1:1:1:1:1). A selected set of codons may also be a set that is biased towards basic amino acids (e.g., His, Lys, Arg). The composition of such a set can be, e.g., AAC : CAC : GAC : TAC : ACC : CCC : GCC : AGC : GGC : TGC : ATC : CTC : GTC : TTC : AAG : CAG : GAG : AGG : TGG : ATG = 1:2:1:1:1:1:1:1:1:1:1:1:2:1:1:2:1:1. Alternatively, a selected set of codons may be a set in which all of the codons code for hydrophobic amino acids (e.g, Pro, Ala, Ile, Leu, Val, Phe, Met) in equal distribution (e.g., CCC : GCC : ATC : CTC : GTC : TTC : ATG = 1:1:1:1:1:1:1).

The nucleosides and nucleotides used herein are often referred to with shorthand designations, in which the protecting group of the 5'-OH is superscripted. For example, "^{T}C" is used to represent N^4-benzoyl 5'-*O*-(4,4'-dimethoxytrityl) 2'-deoxycytidine or N^4-benzoyl 5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxycytidine, and "^{F}G" is used to represent N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl] 2'-deoxyguanosine or N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl] 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyguanosine.

The present invention provides a number of advantages over conventional techniques of nucleic acid synthesis. For example, the methods described herein provide control over codon format (trinucleotide sequences) as well as control over the representation of the codon in a selected set. The invention can therefore generate sets of codons that contain at least one codon for each of the naturally-occurring amino acids and, importantly, can generate

libraries that are substantially free of stop codons. The invention also provides DNA consisting essentially of highly-expressed codons and, as noted above, free of stop codons, which can be efficiently translated to polypeptides of desired lengths. Furthermore, control over codon representation allows for the synthesis of DNA templates that can be used to generate proteins rich in selected amino acids, for example, hydrophobic amino acids, which can be instrumental in protein design techniques.

Brief Description of the Drawings

FIGURES 1, 2, 3, 4, 5, 6, 7, 8, and 9 are each illustrations of coupling sequences for the synthesis of codon libraries.

Description of the Preferred Embodiments

The sequence of bases in DNA and its RNA counterpart determines the sequence of the amino acids in the protein synthesized from this DNA. Sequences of three bases, referred to as codons, correspond to different amino acids. During translation, these codons are read from the 5' end to the 3' end; the resulting protein has an amino acid sequence that corresponds to the sequence of codons.

Three DNA codons, TAA, TAG, and TGA (which correspond to the RNA codons UAA, UAG, and UGA) do not code for any amino acids. Instead, these codons signal release factors to terminate protein synthesis. The presence of stop codons therefore leads to termination of protein synthesis before the entire DNA sequence is translated.

The invention features convenient methods for the controlled synthesis of codon randomized nucleic acids, such as DNA, in which the presence of stop codons can be avoided. If desired, the DNA strand can be used as a template for the synthesis of a complementary DNA strand, which in

turn can serve as a template for the synthesis of the corresponding messenger RNA. Alternatively, messenger RNA can be synthesized directly using the methods of the invention.

As described in more detail below, in the present approach, a desired set of codons, as well as the desired frequency of each codon in the set, is first chosen. The set can include, for example, the most highly-expressed codons for each of the 20 naturally-occurring amino acids, in equal distribution. Highly expressed DNA codons in eukaryotic translation systems typically exhibit either 2'-deoxycytidine (C) or 2'-deoxyguanosine (G) at the 3' end (that is, at the third codon position). Another desired set can include, for example, at least one codon for each of the 20 naturally-occurring amino acids, and in which hydrophobic amino acids are twice as abundant.

In addition, it is generally desirable to omit, from the codon mixture, the known stop codons, TAA, TGA, and TAG, for DNA synthesis, or UAA, UGA, and UAG for RNA synthesis. The omission of these codons from the synthetic library maximizes the likelihood that protein translation is not aborted and that proteins of desired length are generated.

In one particular example, a selected set of highly expressed codons for all 20 naturally-occurring amino acids can be prepared in which all of the codons have a C or G at the third position. Two of the three stop codons, TAA and TGA, are therefore readily excluded from this set.

In addition, none of the stop codons has a C at the third position; codons ending in C can therefore be randomly generated without the introduction of stop codons. The generation of a set of codons ending in C can produce codons for fifteen of the naturally-occurring amino acids.

The exclusion of the stop codon TAG (or, for RNA synthesis, UAG) is more complicated because the best expressed codons for the amino acids Lys, Gln, Glu, Trp, and Met each have a G at position three. The simultaneous

generation of codons for these amino acids, and the exclusion of the stop codon TAG, requires a strategic coupling sequence. Several such coupling sequences are described in detail below as part of the present invention.

The present invention also features methods for generating libraries of codons by using nucleosides and nucleotides with different 5'-protecting groups as building blocks. In preferred methods, the different 5'-protecting groups can be cleaved under orthogonal conditions. In other words, the conditions for cleaving one 5'-protecting group do not cleave the other 5'-protecting groups. An example of one pair of orthogonal protecting groups includes a dimethoxytrityl group (DMT or T), which is cleaved under acidic conditions, and a fluorenylmethyloxycarbonyl group (Fmoc or F), which is cleaved under basic conditions. Another example of a set of orthogonal protecting groups is the set including a dimethoxytrityl group (DMT or T), which is cleaved under acidic conditions, a fluorenylmethyloxycarbonyl group (Fmoc or F), which is cleaved under basic conditions, and a silyl group (S), which is cleaved with fluoride.

In one particular method, a mixture of nucleosides, some of which are protected with a DMT group, and some of which are protected with a Fmoc group, is treated with acid. The DMT-protected nucleosides are deprotected, while the Fmoc protected nucleosides remain protected. When nucleotides are added to this mixture, they couple only with the deprotected nucleosides, allowing for coupling specificity.

In one example, a mixture of N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxycytidine (^{T}C) and N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl]-2'-deoxyguanosine (^{F}G) is treated with acid. The DMT group is cleaved from the C mononucleosides, thus leaving them free to couple with nucleotides. The Fmoc of the G mononucleosides remains attached. Since none of the stop codons end in C, trinucleotides may be randomly generated at this step without

the introduction of stop codons. By this technique, successive coupling steps produce a mixture of trityl-protected trinucleotides ending in C. Throughout the coupling steps, the G mononucleosides remain protected and therefore unreactive.

Codons with G at the third position are then prepared. The Fmoc groups are cleaved with base. The coupling sequence is designed to avoid synthesis of the codon TAG (or UAG).

In yet further preferred embodiments, similar processes are carried out using silyl protecting groups.

Examples of several different coupling schemes are given below. It is to be understood, however, that the invention encompasses additional coupling schemes as well.

Preferably, nucleoside phosphoramidites are used for the coupling reactions. In addition, the internucleotide linkages can be protected with a protecting group, such as an allyl moiety. The allyl protecting group is stable toward both acid and base, but can be cleaved with aqueous ammonia or by palladium (Pd(0)) catalysis.

There now follow particular examples of nucleic acid preparative techniques. These examples are provided for the purpose of illustrating the invention, and should not be construed as limiting.

Example 1A: Synthesis of N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxycytidine (^{T}C)

Allyloxy bis-(diisopropylamino)phosphine is prepared as described in Bannwarth et al., *Tetrahedron Lett.*, 30:4219 (1989), and diisopropylammonium tetrazolide is prepared as described in Barone et al., *Nucleic Acids Res.* 12:4051 (1984). 20 mmol of N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 2'-deoxycytidine and 10 mmol of diisopropylammonium tetrazolide are taken up in anhydrous acetonitrile and evaporated. 25 mmol of

allyloxy bis-(diisopropylamino)phosphine in 200 ml of anhydrous methylene chloride are added to the residue, with stirring. After 10 minutes of stirring at about 25°C, the reaction mixture is poured into 300 ml saturated $NaHCO_3$ solution and extracted with methylene chloride (3 × 200 ml). The combined organic layers are dried over Na_2SO_4 and concentrated. The crude product is purified by chromatography (silica gel, eluting with CH_2Cl_2/MeOH/Et_3N, 94:4:2). The product, N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxycytidine (^{T}C), is precipitated from CH_2Cl_2 into pentane at -60°C.

N^6-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyadenine (^{T}A), N^2-isobutyryl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyguanosine (^{T}G), and 5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxy-thymidine (^{T}T) are prepared using the same reaction conditions.

Example 1B: Alternative synthesis of DMT-allyl dC phosphoramidite monomer

5'-O-DMT-N^4-benzoyl-2'-deoxycytidine (10 g, 15.8 mmol) was dissolved in CH_2Cl_2 (100 mL). Diisopropylammonium tetrazolide (1.35 g, 7.88 mmol) was added followed by allyl-N,N,N,N-tetraisopropylphosphoramidite (5g, 17.3 mmol), and the reaction was stirred overnight at room temperature under argon. The reaction mixture was then extracted with 5% $NaHCO_3$ (3 x 50 mL), H_2O (2 x 50 mL), and dried with Na_2SO_4. The Na_2SO_4 was removed by filtration, and the organics were concentrated to 50 mL under reduced pressure before loading onto a silica gel column (200 g). The column was eluted with EtOAc/heptane/TEA (49/50/1 v:v). Fractions containing the desired product were combined and evaporated under reduced pressure to yield

an oily residue, which was applied to another silica gel column (200 g) and eluted with a stepwise gradient of EtOAc (25 → 75%) in heptane containing 1% TEA. Fractions containing the desired product were combined, concentrated under reduced pressure, dissolved in a small amount of CH_2Cl_2, and precipitated with heptane to give a white powder (8.4 g): 5'-O-Dimethoxytrityl-3'-2-allyl-N,N-diisopropyl-N^4-benzoyl-2'-deoxycytidine phosphoramidite. Purity was determined to be greater than 95% by both RP-HPLC and ^{31}P NMR.

The chemical stability of the new DMT-allyl dC phosphoramidite monomer was monitored by preparing a 0.1M solution in $CDCl_3$ and collecting the ^{31}P NMR spectrum at 24 hour intervals. The monomer was determined to be stable for at least 8 days (i.e., no change in spectrum between 300 and -50 ppm). The coupling ability of the new monomer was evaluated by solid-phase synthesis of the sequence 5'-d(C_9T) on an automated DNA synthesizer (Expedite 8909, PerSeptive Biosystems) using a standard coupling protocol provided by the manufacturer, except that the monomer coupling time was increased to 120 seconds. Trityl absorbance data collected from the instrument indicated that the coupling efficiency was comparable to the same sequence prepared with conventional cyanoethyl phosphoramidites, demonstrating that the DMT-allyl dC monomer couples effectively. After completion of the synthesis, the solid support from the two syntheses (synthesized with either DMT-allyl dC monomer or DMT-cyanoethyl dC monomer) was divided into five portions and treated with 1.5 mL of one of the following at the indicated temperature: concentrated ammonium hydroxide at room temperature, concentrated ammonium hydroxide at 55°C, a mixture of concentrated ammonium hydroxide in ethanol (3:1 v/v) at 55°C, a mixture of t-butyl amine/methanol/water (1:1:2 v/v) at 55°C, 2M anhydrous ammonia in methanol at 55°C. After 24 hours, the room temperature ammonium hydroxide

sample was decanted and concentrated to dryness in a Speed-Vac. For the other four mixtures incubated at 55°C, aliquots were removed 8, 17, and 24 hours, and concentrated to dryness in a Speed-Vac. These 26 samples were then analyzed by anion-exchange HPLC under denaturing conditions (Dionex DNAPac PA-100 column, sodium chloride gradient in 25 mM NaOH, pH 12.4). In the case of 5'-d(C_9T) prepared with DMT-allyl dC phosphoramidite, excellent results were obtained with all conditions except the following: concentrated ammonium hydroxide at room temperature for 24 hours, 2M anhydrous ammonia in methanol at 55°C for 8, 17, or 24 hours. In the case of 5'-d(C_9T) prepared with DMT-cyanoethyl dC phosphoramidite, all of the deprotection reagents completely removed the allyl protecting groups except concentrated ammonium hydroxide at room temperature for 24 hours. The preferred deprotection reagent was determined to be concentrated ammonium hydroxide at 55°C for between 12 and 24 hours.

Example 2A: Synthesis of N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl] 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyguanosine (^{F}G)

3.0 mmol N^2-isobutyryl-2'-deoxyguanosine are co-evaporated from 25 ml pyridine twice, then dissolved in 20 ml pyridine and cooled to 0°C. 3.0 mmol 9-fluorenylmethyl chloroformate (Fmoc-chloride) is added to the stirred solution. The reaction is monitored using thin-layer chromatography (eluting with diethyl ether, then chloroform/methanol 9:1). The reaction is terminated by adding ethanediol; the mixture is then concentrated to an oil. The oil is dissolved in chloroform (150 ml) and washed with saturated $NaHCO_3$ solution. The aqueous phase is extracted twice with chloroform, and the combined chloroform portions are dried over anhydrous Na_2SO_4, filtered, then concentrated to an oil. The oil is co-evaporated from toluene (twice), ethanol, then chloroform, and subjected to a short column chromatography (silica gel)

eluting with a gradient of 0-5% methanol in chloroform. Fractions containing the major product are collected, concentrated to a foam, dissolved in chloroform, precipitated with pentane, filtered, and then dried under vacuum to yield N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl] 2'-deoxyguanosine (^{F}G).

The product is then converted to the phosphoramidite (also referred to as ^{F}G) using the reaction conditions described in Lehmann et al., *Nucleic Acids Res.*, Vol. 17, No. 7, 2379-2390 (1989).

N^6-benzoyl-5'-*O*-[(9-fluorenyl)methoxycarbonyl] 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyadenine (^{F}A) is prepared using the same reaction conditions.

Example 2B: Alternative synthesis of Fmoc-allyl dG phosphoramidite monomer and its application in DNA synthesis

N^2-Isobutyryl-2'-deoxyguanosine (15 g, 44.5 mmol) was evaporated from pyridine (3 x 100 mL) and dissolved in anhydrous pyridine (150 mL). The solution was cooled to 0°C and Fmoc-Cl (12.6 g, 49 mmol) was added. After completion of the reaction, as indicated by TLC, the mixture was evaporated to dryness, redissolved in CH_2Cl_2 (200 mL) and washed with 5% $NaHCO_3$ (2 x 75 mL) followed by H_2O (2 x 75 mL) and brine (1 x 75 mL). The organic layer was dried with Na_2SO_4, filtered, and concentrated to a low volume under reduced pressure. The concentrated organic layer was applied to a silica gel column (700 g) and eluted first with EtOAc, and then with MeOH/CH_2Cl_2/EtOAc (5/30/65 v:v). Fractions containing the desired product were combined and concentrated to dryness under reduced pressure to yield a slightly yellow solid (7.1 g).

5'-O-Fmoc-N^2-isobutyryldeoxyguanosine (8 g, 14.3 mmol) prepared above was evaporated from pyridine (2 x 100 mL) and then acetonitrile (3 x

100 mL). The resulting solid was dissolved in CH_2Cl_2 (100 mL). Allyl-N,N,N,N-tetraisopropylphosphoramidite (5.45 mL, 17.2 mmol) was added to the reaction mixture followed by diisopropylammonium tetrazolide (0.62 g, 3.6 mmol). After two hours of stirring at room temperature, the reaction mixture was washed with 5% $NaHCO_3$ (1 x 50 mL), H_2O (1 x 100 mL), and then dried with Na_2SO_4, filtered, and concentrated under reduced pressure. The concentrated mixture was applied to a silica gel column (500 g) and eluted with a stepwise gradient of EtOAc (25 → 95%) in heptane containing 1% lutidine, according to the procedure described in Lehmann et al., Nucleic Acids Research 17: 2379 (1989). Fractions containing the desired product were combined and concentrated to dryness under reduced pressure. The residue was dissolved in toluene (15 mL) and precipitated into stirred heptane (1 L). Filtration yielded an off-white powder (4.5 g) which was further purified by silica gel chromatography. The silica gel (400 g) was packed with EtOAc/heptane/lutidine (79/19/2 v:v) and then washed with EtOAc/heptane (80/20 v:v) prior to applying the partially purified material (4.2 g) from above. The column was eluted with EtOAc/heptane (80/20 v:v), and fractions containing the desired product were combined and evaporated under reduced pressure in the presence of anhydrous toluene (3 x 20 mL) followed by evaporation from anhydrous acetonitrile (2 x 30 mL). Only the last evaporation was taken to dryness, which yielded a white foam (2.6 g): 5'-O-Fmoc-3'-2-allyl-N,N-diisopropyl-N^2-isobutyryl-2'-deoxyguanosine phosphoramidite. Purity and identity were established by RP-HPLC, ^{31}P NMR and ^{1}H NMR.

The chemical stability of the new Fmoc-allyl dG phosphoramidite monomer was monitored by preparing a 0.1M solution in $CDCl_3$ and collecting the ^{31}P NMR spectrum at 24 hour intervals. The monomer was 10% degraded after 1 day and 50% degraded after 3 days, as indicated by the appearance and

growth of a new peak at 153 ppm resulting from spontaneous loss of the 5'-Fmoc group. As most syntheses are completed within several hours, the stability of the Fmoc-allyl dG phosphoramidite was deemed suitable. The coupling ability of the new monomer was evaluated by solid-phase synthesis of the sequence 5'-d(G_9T) on an automated DNA synthesizer (Expedite 8909, PerSeptive Biosystems). The standard synthesis protocol provided by the manufacturer was modified to increase the coupling time (900 sec.), increase the capping step (120 sec.), increase the oxidation time (60 sec.), and deliver the 5'-Fmoc deprotection reagent for 120 seconds from an auxiliary bottle position. Both 0.1M DBU in acetonitrile and 0.1M piperidine in anhydrous DMF were evaluated as 5'-Fmoc deprotection reagents. The completed 5'-d(G_9T) sequences were deprotected in concentrated ammonium hydroxide for 18 hours at 55°C, concentrated in a Speed-Vac, analyzed by anion-exchange HPLC under denaturing conditions (Dionex DNAPac PA-100 column, sodium chloride gradient in 25 mM NaOH, pH 12.4), and compared to a control sequence synthesized with standard DMT-dG cyanoethyl phosphoramidites. 0.1M Piperidine in DMF was the preferred 5'-Fmoc deprotection reagent.

Example 3A: Synthesis of N^2-isobutyryl-5'-*O*-[trimethylsilyl] 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyguanosine (^{S}G)

2.0 mmol N^2-isobutyryl-2'-deoxyguanosine are dissolved in 20 ml DMF and stirred at 25°C. 3.0 mmol trimethylsilyl chloride and 0.5 mmol imidazole are added to the stirred solution. The reaction is monitored using thin-layer chromatography. When the reaction is complete, the mixture is concentrated to an oil. The oil is dissolved in chloroform and washed with saturated $NaHCO_3$ solution. The aqueous phase is extracted twice with chloroform, and the combined chloroform portions are dried over anhydrous

Na_2SO_4, filtered, then concentrated to an oil. The oil is co-evaporated from toluene (twice), ethanol, then chloroform, and subjected to a short column chromatography (silica gel) eluting with a gradient of 0-5% methanol in chloroform. Fractions containing the major product are collected, concentrated to a foam, dissolved in chloroform, precipitated with pentane, filtered, and then dried under vacuum to yield N^2-isobutyryl-5'-*O*-[trimethylsilyl] 2'-deoxyguanosine (^{S}G).

The product is then converted to the phosphoramidite (also referred to as ^{S}G) using the reaction conditions described in Example 1.

Example 3B: Synthesis of silyl-allyl dG phosphoramidite monomer and its application in DNA synthesis

N^2-Isobutyryl-2'-deoxyguanosine (6.75 g, 20 mmol) was evaporated from pyridine (3 x 100 mL), dissolved in anhydrous pyridine (75 mL) and cooled to 0°C. Bis(trimethylsiloxy)cyclododecyloxy-silyl chloride (8.5 g, 22 mmol) was added to the stirred solution. After two hours the reaction mixture was concentrated to dryness under reduced pressure and resuspended in CH_2Cl_2 (100 mL). This solution was washed with 5% $NaHCO_3$ (2 x 30 mL), H_2O (2 x 30 mL), and then dried with Na_2SO_4, filtered, and concentrated under reduced pressure. The concentrated mixture was applied to a silica gel column (500 g) and eluted with a stepwise gradient of MeOH (0 → 10%) in CH_2Cl_2. Fractions containing the desired product were combined and concentrated to dryness under reduced pressure to yield a white solid (9.5 g).

5'-O-Bis(trimethylsiloxy)cyclododecyloxy-silyl-N^2-isobutyryl-2'-deoxyguanosine (9 g, 12.3 mmol) from above was evaporated first from pyridine (2 x 100 mL) and then acetonitrile (3 x 100 mL). The residue was dissolved in anhydrous CH_2Cl_2 (100 mL) and allyl-N,N,N,N-tetraisopropylphosphoramidite (4.2 mL, 14.5 mmol) was added

to the stirred reaction mixture followed by diisopropylammonium tetrazolide (0.56 g, 3.0 mmol). After two hours the reaction mixture was washed with 5% $NaHCO_3$ (1 x 50 mL), H_2O (1 x 100 mL) and then dried with Na_2SO_4, filtered, and concentrated under reduced pressure. The concentrated solution was applied to a silica gel column (500 g) that had been packed with EtOAc/hexane (70:30 v:v) containing 5% TEA. The product was eluted with EtOAc/hexane (70:30 v:v) containing 2% TEA. Fractions containing the desired product were combined and concentrated to dryness under reduced pressure. The residue was taken up in toluene (100 mL) and evaporated to dryness two times, and this process was repeated with anhydrous acetonitrile (3 x 100 mL) to finally give a white foam (8.1 g): 5'-O-Bis(trimethylsiloxy)cyclododecyloxy-silyl-3'-2-allyl-N,N-diisopropyl-N^2-isobutyryl-2'-deoxyguanosine phosphoramidite. Purity was determined to be greater than 98% by both ^{31}P NMR and RP-HPLC.

The coupling ability of the new Silyl-allyl dG monomer was evaluated by solid-phase synthesis of the sequence 5'-d(G_9T) on an automated DNA synthesizer (Expedite 8909, PerSeptive Biosystems) using a polystyrene solid support (PE BioSystems, Foster City, CA). The standard 0.2 μmole cyanoethyl phosphoramidite synthesis protocol provided by the manufacturer was modified to accommodate the new chemistries. The modified protocol contained longer monomer coupling steps (240 sec.), longer wash times (120 sec.), and new cycles to deliver the non-standard Silyl deprotection reagent (HF/TEA, 1.1M:1.6M in DMF) from an auxiliary bottle position. In addition, the standard trichloroacetic acid reagent was replaced with 3% dichloroacetic acid in CH_2Cl_2. The completed 5'-d(G_9T) sequences were deprotected in concentrated ammonium hydroxide for 18 hours at 55°C, concentrated in a Speed-Vac, analyzed by anion-exchange HPLC under denaturing conditions (Dionex DNAPac PA-100 column, sodium chloride gradient in 25 mM NaOH, pH 12.4), and compared to a control sequence synthesized with standard

DMT-dG cyanoethyl phosphoramidites. These materials were also analyzed by MALDI-TOF mass spectrometry and gave the expected signal at m/z 3206.07 for the sequence prepared with the new Silyl-allyl dG phosphoramidite monomer and m/z 3206.36 for the sequence prepared with conventional DMT-cyanoethyl dG phosphoramidite monomer (theoretical mass of $d(G_9T)$ = 3205.12).

Example 4: Synthesis of 3'-*O*-*tert*-butyl-dimethylsilyl 2-deoxythymidine

20 mmol 5'-*O*-(4,4'-dimethoxytrityl) 2'-deoxythymidine are dissolved in 200 ml DMF and stirred at 25°C. 30 mmol *tert*-butyldimethylsilyl chloride and 5 mmol imidazole are added to the stirred solution. The reaction is monitored using thin-layer chromatography. When the reaction is complete, the mixture is concentrated to an oil. The oil is dissolved in chloroform (150 ml) and washed with saturated $NaHCO_3$ solution. The aqueous phase is extracted twice with chloroform; the combined chloroform portions are dried over anhydrous Na_2SO_4, filtered, then concentrated to an oil. The oil is co-evaporated from toluene (twice), ethanol, then chloroform, and subjected to a short column chromatography (silica gel) eluting with a gradient of 0-5% methanol in chloroform. Fractions containing the major product are collected, concentrated to a foam, dissolved in chloroform, precipitated with pentane, filtered, and then dried under vacuum.

The DMT protecting group is then cleaved as follows. The product is dissolved in 75 ml CH_2Cl_2/MeOH (8:1 v/v); Amberlyst® 15 ion exchange resin is then added in portions until the surface of the resin remains orange colored. The suspension is stirred 24 hours, the resin is filtered off, and the solution is concentrated *in vacuo*. The product is precipitated twice from petroleum ether (500 ml) at 40-60°C.

N^6-benzoyl-3'-*O*-*tert*-butyl-dimethylsilyl 2'-deoxyadenine, N^2-isobutyryl-3'-*O*-*tert*-butyl-dimethylsilyl 2'-deoxyguanosine, and N^4-benzoyl-3'-*O*-*tert*-butyl-dimethylsilyl 2'-deoxycytidine are prepared using the same reaction conditions.

Example 5: Synthesis of ^{T}AT dinucleotide phosphoroamidite

A solution containing a mixture of 15 mmol 3'-*O*-*tert*-butyl-dimethylsilyl 2'-deoxythymidine and 24 mmol tetrazole is dried by repeated coevaporation with acetonitrile/toluene. The mixture is then dissolved in 50 ml dry acetonitrile. 15 mmol N^6-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyadenine, which is pre-dried by repeated coevaporation with toluene, in 30 ml dry acetonitrile is added. The reaction is followed by TLC. If the reaction does not go to completion, additional phosphoramidite can be added. When the reaction is complete, the reaction mixture is cooled in an ice bath, and 40 mmol *tert*-butyl hydroperoxide is added. After about 15 minutes, the solution is concentrated *in vacuo*. The oil is dissolved in ethyl acetate and washed with a phosphate buffer (pH = 6.8) and water. The solution is dried over anhydrous Na_2SO_4, filtered, and concentrated to dryness.

The TBDMS ether protecting group is cleaved as follows. The product is dissolved in 40 ml THF. 30 mmol tetrabutylammonium fluoride is added, and the reaction mixture is stirred 1 hour at 25 °C. The THF is evaporated *in vacuo*; water is then added to the concentrated reaction mixture. The resulting mixture is extracted with CH_2Cl_2 (3 ×100 ml). The combined organic layers are dried over Na_2SO_4, filtered, and concentrated. The product is then purified with column chromatography (silica gel, using methanol in CH_2Cl_2 to elute). The product is then converted to the phosphoramidite using the reaction conditions described in Example 1.

Other dinucleotide phosphoramidites (e.g., ^{T}TG and ^{T}AT) are prepared using the same reaction conditions.

Example 6: Synthesis of N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxycytidine 3'-*O*-succinic acid

N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxycytidine and succinic anhydride (10-fold excess) are dissolved in DMF and stirred at 70°C for 40 hours. The reaction is monitored by TLC (silica gel, development in ether, then chloroform/methanol 9:1). After completion of the reaction, the reaction mixture is taken up in methylene chloride, then washed with 20% aqueous citric acid solution. The aqueous phase is washed twice with methylene chloride. The combined organic layers are washed with water, dried over anhydrous Na_2SO_4, and concentrated to dryness. The product is purified by chromatography (silica gel, eluting with chloroform and chloroform/methanol 99:1).

N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl]-2'-deoxyguanosine 3'-*O*-succinic acid is prepared using the same reaction conditions.

Example 7: Functionalization of support

A glass support for use in DNA synthesis is treated with Fmoc-sarcosine in the presence of dicyclohexylcarbodiimide, followed by removed of the Fmoc group with piperidine/DMF.

N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxycytidine 3'-*O*-succinic acid and N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl]-2'-deoxyguanosine 3'-*O*-succinic acid are dissolved in THF; dicyclohexylcarbodiimide is added to the solution. The reaction is stirred for 0.5 hours at 25°C, then filtered. The filtrate is evaporated to dryness. The

residue is dissolved in DMF, filtered, and shaken with the functionalized glass support for 16 hours. The support is separated by filtration, washed with methylene chloride and diethyl ether. Unreacted amino groups are capped by treatment of the support with a mixture of THF/lutidine/acetic anhydride (8:1:1) and *N*-methylimidazole in THF. The support is then washed with methylene chloride and diethyl ether, and dried *in vacuo*.

Example 8A: Synthesis of codons

In one preferred synthetic approach, the codons are built up from the 3'-end, as shown in Figure 1, using solid phase synthesis. A solid phase synthesizer is used, according to the manufacturer's instructions.

A 16:5 mixture of N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxycytidine (^{T}C) and N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl]-2'-deoxyguanosine (^{F}G) is attached to a support, as described in Examples 6 and 7.

After the nucleosides have been attached to the glass support, trichloroacetic acid is added to cleave the trityl protecting groups from the ^{T}C mononucleosides. Since the Fmoc protecting group is not labile under acidic conditions, the ^{F}G mononucleosides remain protected, and therefore unreactive.

A 1:1:1:1 mixture of N^6-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyadenine (^{T}A), N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxycytidine (^{T}C), N^2-isobutyryl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyguanosine (^{T}G), and 5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxythymidine (^{T}T) mononucleoside phosphoramidites is pre-dried by repeated coevaporation with acetonitrile/toluene, then dissolved in dry acetonitrile. The mixture is then added to the mixture of C and ^{F}G mononucleosides. When the coupling reaction is complete, a solution of 0.02

iodine in THF/pyridine/water is added to oxidize the products. The result of this series of reactions is a 1:1:1:1 mixture of ^{T}AC, ^{T}CC, ^{T}GC, and ^{T}TC dinucleotides, and ^{F}G mononucleoside.

The trityl protecting groups of the dinucleotides are then cleaved with acid. The dinucleotides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result is a mixture of 16 unique codons, each corresponding to a different amino acid (with the exception of TTC and AGC, which both represent serine), and ^{F}G mononucleosides.

The Fmoc protecting groups of the G mononucleosides are then cleaved with 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU), as described in Lehmann et al., *Nucleic Acids Res.* 17:2379 (1989). The trityl protecting groups of the trinucleotides are not labile under basic conditions; the trinucleotides therefore remain unreactive. The deprotected G mononucleosides are coupled with a 3:1:1 mixture of ^{F}A mononucleoside phosphoramidites, ^{T}TG dinucleotide phosphoramidite, and ^{T}AT dinucleotide phosphoramidite, and the products of the coupling reactions are oxidized. The result is two more trinucleotide codons, and ^{F}AG dinucleotides.

The Fmoc protecting groups of the dinucleotides are once again cleaved with base, the dinucleotides are coupled with a 1:1:1 mixture of ^{T}A, ^{T}C, and ^{T}G mononucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The nucleotide T is omitted, as the inclusion of this nucleotide at this point would result in the synthesis of a TAG codon.

As shown in Figure 1, the end result of the successive deprotection and coupling reactions is a mixture of 21 codons, each corresponding to one of the 20 naturally occurring amino acids. All 20 amino acids are represented, and only one amino acid is represented twice. The invention therefore provides a synthesis of a codon set in which all of the amino acids are represented

approximately equally. Most importantly, the set contains substantially no stop codons.

The trityl group can be removed from the trinucleotides, and a mixture of ^{T}C and ^{F}G nucleoside phosphoramidites can be added. The process for synthesizing the codons can then be repeated until DNA of the desired length is achieved.

Example 8B: Synthesis of CCC/CGC codons (Pro/Arg) via acid/base orthogonal deprotection

The tetramer 5'-d(CSCT), where S is either G or C, was synthesized on an automated DNA synthesizer (Expedite 8909, PerSeptive Biosystems) using the acid/base orthogonal deprotection scheme employing both DMT and Fmoc 5'-hydroxyl protecting groups with allyl-P protection. The standard 0.2 μmole cyanoethyl phosphoramidite synthesis protocol provided by the manufacturer was modified to accommodate the new chemistries. The modified protocol contained longer monomer coupling steps (240 sec.), longer wash times (120 sec.), and new cycles to deliver the non-standard Fmoc deprotection reagent (0.1M piperidine in anhydrous DMF). In addition, the standard trichloroacetic acid reagent was replaced with 3% dichlorocetic acid in CH_2Cl_2. CPG solid support functionalized with 0.2 μmole T monomer was loaded onto the instrument, and DMT-C monomer was added according to the modified protocol. After removal of the 5'-DMT group, equal volumes of Fmoc-G and DMT-C monomer were delivered to the column with the tetrazole coupling agent to form a mixture of two trimers on the solid support: GCT (with 5'-Fmoc protection), and CCT (with 5'-DMT protection). The 5'-DMT was removed with 3% dichloroacetic acid in CH_2Cl_2 and DMT-C monomer was then delivered to the column to extend the CCT sequence to CCCT. Next, a 0.1M piperidine solution in DMF was delivered to the column to remove the

5'-Fmoc protecting group. DMT-C monomer was again added to the column to form CGCT from the remaining GCT sequence. The terminal 5'-DMT was removed with 3% dichloroacetic acid in CH_2Cl_2 and the CPG support was treated with concentrated ammonium hydroxide at 55°C for 16 hours. The solution was finally cooled, concentrated to dryness on a Speed-Vac, and taken up in water. This material was analyzed by MALDI-TOF mass spectrometry and gave the expected signals at m/z 1110.84 (for dCCCT; theoretical = 1110.80) and m/z 1151.23 (for dCGCT; theoretical = 1150.82). A small amount of the crude material was also degraded enzymatically with a mixture of snake venom phosphodiesterase and bacterial alkaline phosphatase to establish the nucleoside ratio. The resulting digest was analyzed quantitatively by RP-HPLC according to the general scheme described in Eadie et al., Anal. Biochem. 165: 442 (1987). The tetramer standard (prepared with conventional cyanoethyl phosphoramidites and the manufacturer's S coupling cycle at base position three) produced 2.4:0.6:1.0 for the normalized ratio of C:G:T nucleosides, respectively, compared to a theoretical value of 2.5:0.5:1. The same tetramer prepared via the acid/base orthogonal deprotection scheme produced 2.2:0.8:1.0 for the normalized ratio of C:G:T nucleosides.

The 15-mer 5'-d(ACGTGGCTGAACSCT), where S is either G or C, was also synthesized on an automated DNA synthesizer using the same acid/base orthogonal deprotection scheme described above. The terminal 5'-DMT was removed with 3% dichloroacetic acid in CH_2Cl_2, and the CPG support was treated with concentrated ammonium hydroxide at 55°C for 16 hours. The solution was finally cooled, concentrated to dryness on a Speed-Vac, and taken up in water. The mixture was analyzed by anion-exchange HPLC under denaturing conditions (Dionex DNAPac PA-100 column, sodium chloride gradient in 25 mM NaOH, pH 12.4) and revealed two closely spaced peaks (retention time difference = 30 sec.), corresponding to the

two 15-mers differing by one base (C or G at position 3). In the case of the sequences prepared via the acid/base orthogonal deprotection scheme, the ratio of the peak areas was 0.61:0.39, whereas in the case of the standard (prepared with conventional cyanoethyl phosphoramidites and the manufacturer's S coupling cycle at base position three) the ratio of the peak areas was 0.51:0.49. The sequence prepared by the orthogonal deprotection scheme was also analyzed by MALDI-TOF mass spectrometry and gave the expected signals at m/z 4553.63 (for dACGTGGCTGAACCCT; theoretical = 4553.04) and m/z 4592.56 (for dACGTGGCTGAACGCT; theoretical = 4593.07).

Example 8C: Synthesis of CCC/CGC codons (Pro/Arg) via acid/fluoride orthogonal deprotection

The tetramer 5'-d(CSCT), where S is either G or C, was synthesized on an automated DNA synthesizer using the acid/fluoride orthogonal deprotection scheme employing both DMT and Silyl 5'-hydroxyl protecting groups with allyl-P protection. The standard 0.2 μmole cyanoethyl phosphoramidite synthesis protocol provided by the manufacturer was modified to accommodate the new chemistries. The modified protocol contained longer monomer coupling steps (240 sec.), longer wash times (120 sec.), and new cycles to deliver the non-standard silyl deprotection reagent (HF/TEA, 1.1M:1.6M in DMF) over 180 seconds. In addition, the standard trichloroacetic acid reagent was replaced with 3% dichlorocetic acid in CH_2Cl_2. Polystyrene solid support functionalized with 0.2 μmole T monomer (PE Biosystems, Foster City, CA) was loaded onto the instrument, and DMT-C monomer was added according to the modified protocol. After removal of the 5'-DMT group, equal volumes of Silyl-G and DMT-C were delivered to the column with the tetrazole coupling agent to form a mixture of two trimers: GCT (with 5'-Silyl protection), and CCT (with 5'-DMT protection). The 5'-DMT was removed

with 3% dichloroacetic acid in CH_2Cl_2 and DMT-C monomer was delivered to the column to extend the CCT sequence to CCCT. Next, an HF/TEA mixture in DMF (1.1M:1.6M) was delivered to the column to remove the 5'-Silyl protecting group. DMT-C monomer was again added to the column to form CGCT from the remaining GCT sequence. The terminal 5'-DMT was removed with 3% dichloroacetic acid in CH_2Cl_2 and the CPG support was treated with concentrated ammonium hydroxide at 55°C for eight hours. The solution was finally cooled and concentrated on a Speed-Vac. A portion of this material was purified by anion-exchange HPLC (Dionex DNAPac PA-100 column, sodium chloride gradient in 25 mM NaOH, pH 12.4), and degraded enzymatically with a mixture of snake venom phosphodiesterase and bacterial alkaline phosphatase to establish the nucleoside ratio. The resulting digest was analyzed quantitatively by RP-HPLC according to the general scheme described in Eadie et al., Anal. Biochem. 165: 442 (1987). The tetramer standard (prepared with conventional cyanoethyl phosphoramidites and the manufacturer's S coupling cycle at base position three) produced 2.0:1.0:1.0 for the normalized ratio of C:G:T nucleosides, respectively, compared to a theoretical value of 2.5:0.5:1. The same tetramer prepared via the acid/base orthogonal deprotection scheme produced the normalized ratio 1.9:1.1:1.0 for C:G:T nucleosides.

The 15-mer 5'-d(ACGTGGCTGAACSCT), where S is either G or C, was synthesized on an automated DNA synthesizer using the acid/fluoride orthogonal deprotection scheme described above. The terminal 5'-DMT was removed with 3% dichloroacetic acid in CH_2Cl_2 and the polystyrene support was treated with concentrated ammonium hydroxide at 55°C for 16 hours. The solution was finally cooled, concentrated to dryness on a Speed-Vac, and taken up in water. The mixture was analyzed by anion-exchange HPLC under denaturing conditions (Dionex DNAPac PA-100 column, sodium chloride gradient in 25 mM NaOH, pH 12.4) and revealed two closely spaced peaks

(retention time difference = 30 sec.) corresponding to the two 15-mers differing by one base (C vs. G) at position three. In the case of the sequences prepared via the acid/fluoride orthogonal deprotection scheme, the ratio of the peak areas at 260 nm was 0.34:0.66, whereas in the case of the standard (prepared with conventional cyanoethyl phosphoramidites and the manufacturer's S coupling cycle at base position three) the ratio of the peak areas was 0.4:0.6. This material was also analyzed by MALDI-TOF mass spectrometry and gave the expected signals at m/z 455267.63 (for dACGTGGCTGAACCCT; theoretical = 4553.04) and m/z 4593.41 (for dACGTGGCTGAACGCT; theoretical = 4593.07).

Example 9: Removal of oligonucleotide from support

At the end of the above-described coupling reactions, the support is treated with concentrated ammonia at 70°C for 2 hours in a tightly closed Eppendorf tube, to cleave the oligonucleotides from the support. After filtration, the ammonia solution is evaporated on a speed-vac concentrator. The residue is taken up in water and centrifuged (15 minutes, 0°C). DNA is precipitated from the supernatant by the addition of dioxane and THF. After centrifuging (15 minutes, 0°C), the pellet is dissolved in water. The product DNA is purified by reverse-phase HPLC.

Alternatively, the support material is treated under argon with $Pd(PPh_3)_4$/morpholine in THF/DMSO/0.5 M HCl (2/2/2/1) at 25°C. The support is washed with THF and acetone and treated with concentrated NH_3 for 2 hours at 25°C. After filtration the ammonia solution is evaporated, the residue is dissolved in water, and the DNA is purified by HPLC.

Example 10: Synthesis of random codons

In other preferred synthetic approaches, examples 10-13 are carried

out using the general methods described in Example 8; the successive coupling reactions take place in the same reaction vessel.

In a first approach, a 14:6 mixture of ^{T}C and ^{F}G is attached to a support, as described in Examples 6 and 7. The trityl protecting groups are cleaved with trichloroacetic acid. The C mononucleosides are then coupled with a 1:1:1:1:1:1:1:1:1:1:1:1:1:1 mixture of ^{T}AA, ^{T}CA, ^{T}GA, ^{T}TA, ^{T}AC, ^{T}CC, ^{T}GC, ^{T}AG, ^{T}GG, ^{T}TG, ^{T}AT, ^{T}CT, ^{T}GT, and ^{T}TT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized. The result of these reactions is a mixture of 14 unique codons, each representing a different amino acid, and ^{F}G mononucleoside, as shown in Figure 2.

The Fmoc protecting groups of the G mononucleosides are then cleaved with DBU, as described in Example 6. The deprotected mononucleosides are coupled with a 1:1:1:1:1:1 mixture of ^{T}AA, ^{T}CA, ^{T}GA, ^{T}AG, ^{T}TG, and ^{T}AT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized. The end result of these coupling reactions is a mixture of 20 unique trinucleotides, each representing a codon for one of the 20 naturally-occurring amino acids, as shown in Figure 2. Once again, no stop codons are present in the mixture.

This process for synthesizing the codons can be repeated until DNA of the desired length is achieved.

Example 11: Synthesis of codons

In another preferred approach, a 16:5 mixture of ^{T}C and ^{F}G mononucleosides is attached to a support. The trityl protecting groups are cleaved with trichloroacetic acid.

The C mononucleosides are then coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result of these reactions is a 1:1:1:1

mixture of ^{T}AC, ^{T}CC, ^{T}GC, and ^{T}TC dinucleotides, and ^{F}G mononucleoside.

The trityl protecting groups of the dinucleotides are then cleaved with acid. The dinucleotides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result is a mixture of 16 unique codons, each representing a different amino acid (with the exception of TTC and AGC, which both correspond to serine), and ^{F}G mononucleoside.

The protecting groups of the ^{F}G mononucleosides are then cleaved with DBU. The G mononucleosides are coupled with a 1:1:1:1:1 mixture of ^{T}AA, ^{T}CA, ^{T}GA, ^{T}TG, and ^{T}AT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized.

As shown in Figure 3, the end result of the successive deprotection and coupling reactions is a mixture of 21 codons. The process for synthesizing the codons can be repeated until DNA of the desired length is achieved.

Example 12: Synthesis of trinucleotides

In yet another preferred synthetic approach, a 16:6 mixture of ^{T}C and ^{F}G mononucleosides is attached to a support, and the trityl protecting groups are cleaved with trichloroacetic acid. The C mononucleosides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result of these reactions is a 1:1:1:1 mixture of ^{T}AC, ^{T}CC, ^{T}GC, and ^{T}TC dinucleotides, and ^{F}G mononucleoside.

The trityl protecting groups of the dinucleotides are then cleaved with acid. The dinucleotides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T mononucleoside phosphoramidites, and the products of the coupling reactions are oxidized.

The protecting groups of the ^{F}G mononucleosides are then cleaved

with DBU. The G mononucleosides are coupled with a 3:1:1:1 mixture of ^{F}A nucleoside phosphoramidite and ^{T}TG, ^{T}AT, and ^{T}CU dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized. The Fmoc protecting groups are cleaved from the dinucleotides, and a 1:1:1 mixture of ^{T}A, ^{T}C, and ^{T}G mononucleoside phosphoramidites is added; the products of the coupling reactions are then oxidized.

As shown in Figure 4, the result of these successive deprotection and coupling reactions is a mixture of 22 codons, each corresponding to an amino acid. The synthetic scheme results in the generation of a set of codons in which the amino acids Ser and Leu are twice as abundant as the other naturally occurring amino acids. This distribution is close to the amino acid distribution typically found in biological proteins. The process for synthesizing the codons can be repeated until DNA of the desired length is achieved.

Example 13: Synthesis of trinucleotides

In another preferred synthetic approach, a 16:6 mixture of ^{T}C and ^{F}G mononucleotides is attached to a support. The trityl protecting groups are cleaved with trichloroacetic acid.

The C mononucleosides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result of these reactions is a 1:1:1:1 mixture of ^{T}AC, ^{T}CC, ^{T}GC, and ^{T}TC dinucleotides, and ^{F}G mononucleoside.

The trityl protecting groups of the dinucleotides are then cleaved with acid. The dinucleotides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized.

The protecting groups of the ^{F}G mononucleotides are then cleaved with DBU. The G mononucleosides are coupled with a 1:1:1:1:1:1 mixture of

^{T}AA, ^{T}CA, ^{T}GA, ^{T}TG, ^{T}AT, and ^{T}CU dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized.

As shown in Figure 5, the result of the successive deprotection and coupling reactions is a mixture of 22 codons. The synthetic scheme results in the generation of a set of codons in which the amino acids Ser and Leu are twice as abundant as the other naturally occurring amino acids. This distribution represents the amino acid distribution found in biological proteins. The process for synthesizing the codons can be repeated until DNA of the desired length is achieved.

Example 14: Synthesis of trinucleotides using three protecting groups

In an additional preferred synthetic approach, a 16:3:2 mixture of ^{T}C, ^{F}G, and ^{S}G mononucleosides is attached to a support. The trityl protecting groups are cleaved with trichloroacetic acid.

The C mononucleosides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result of these reactions is a 1:1:1:1 mixture of ^{T}AC, ^{T}CC, ^{T}GC, and ^{T}TC dinucleotides, ^{F}G mononucleosides, and ^{S}G mononucleosides.

The trityl protecting groups of the dinucleotides are cleaved with acid. The dinucleotides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized.

The protecting groups of the ^{F}G mononucleotides are cleaved with DBU. The G mononucleosides are coupled with ^{F}A mononucleoside phosphoramidite, and the products of the coupling reactions are oxidized.

The Fmoc protecting groups are again cleaved. The dinucleotides

are coupled with a 1:1:1 mixture of ^{T}A, ^{T}C, and ^{T}G mononucleoside phosphoramidites, and the products of the coupling reactions are oxidized.

The silyl protecting groups are cleaved with anhydrous tetra-*n*-butylammonium fluoride. The G mononucleosides are coupled with a 1:1 mixture of ^{F}G and ^{S}T mononucleoside phosphoramidites, and the products of the coupling reactions are oxidized.

The Fmoc protecting group of the dinucleotide is cleaved, and the dinucleotide is coupled with ^{T}T mononucleoside phosphoramidite. The product of the coupling reaction is oxidized.

Finally, the silyl group of the ^{S}TG dinucleotide is cleaved. The dinucleotide is coupled with ^{T}A, and the product is oxidized.

As shown in Figure 6, the result of the successive deprotection and coupling reactions is a mixture of 21 codons. The process for synthesizing the codons can be repeated until DNA of the desired length is achieved.

Example 15: Synthesis of hydrophobic amino acids

In yet another preferred synthetic approach, a 6:1 mixture of ^{T}C and ^{F}G mononucleosides is attached to a support, and the trityl protecting groups are cleaved with trichloroacetic acid. The C mononucleosides are coupled with a 1:1:1:1:1:1 mixture of ^{T}CC, ^{T}GC, ^{T}AT, ^{T}CT, ^{T}GT, and ^{T}TT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized.

The Fmoc protecting group of the ^{F}G mononucleoside is then cleaved. The mononucleoside is coupled with a ^{T}AT dinucleotide phosphoramidite, and the product of the coupling reaction is oxidized.

As shown in Figure 7, the result of these successive deprotection and coupling reactions is a mixture of 7 codons, each corresponding to a hydrophobic amino acid (Pro, Ala, Ile, Leu, Val, Phe, or Met).

Example 16: Synthesis of codons with a bias for hydrophobic amino acids

In another preferred approach, a 16:5 mixture of ^{T}C and ^{F}G mononucleosides is attached to a support. The trityl protecting groups are cleaved with trichloroacetic acid.

The C mononucleosides are then coupled with a 1:1:1:2 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result of these reactions is a 1:1:1:2 mixture of ^{T}AC, ^{T}CC, ^{T}GC, and ^{T}TC dinucleotides, and ^{F}G mononucleoside.

The trityl protecting groups of the dinucleotides are then cleaved with acid. The dinucleotides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized.

The protecting groups of the ^{F}G mononucleosides are then cleaved with DBU. The G mononucleosides are coupled with a 1:1:1:1:1 mixture of ^{T}AA, ^{T}CA, ^{T}GA, ^{T}TG, and ^{T}AT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized.

As shown in Figure 8, the end result of the successive deprotection and coupling reactions is a mixture of 20 codons; the codons ATC, CTG, GTC, and TTC, which correspond to the hydrophobic amino acids Ile, Leu, Val, and Phe, are represented twice. The process for synthesizing the codons can be repeated until DNA of the desired length is achieved. The resulting DNA will code for proteins with a high percentage of hydrophobic amino acids.

Example 17: Synthesis of codons with a bias for basic amino acids

In another preferred approach, a 14:6 mixture of ^{T}C and ^{F}G mononucleosides is attached to a support. The trityl protecting groups are cleaved with trichloroacetic acid.

The C mononucleosides are then coupled with a 1:2:1:1:1:1:1:1:1:1:1:1:1:1:1:1 mixture of ^{T}AA, ^{T}CA, ^{T}GA, ^{T}TA, ^{T}AC, ^{T}CC, ^{T}GC, ^{T}AG, ^{T}GG, ^{T}TG, ^{T}AT, ^{T}CT, ^{T}GT, and ^{T}TT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized.

The protecting groups of the ^{F}G mononucleosides are then cleaved with DBU. The G mononucleosides are coupled with a 2:1:1:2:1:1 mixture of ^{T}AA, ^{T}CA, ^{T}GA, ^{T}AG, ^{T}TG, and ^{T}AT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized.

As shown in Figure 9, the end result of the successive deprotection and coupling reactions is a mixture of 20 codons; the codons CAC, AAG, and AGG, which correspond to the basic amino acids His, Lys, and Arg, are represented twice. The process for synthesizing the codons can be repeated until DNA of the desired length is achieved. The resulting DNA will code for proteins with a high percentage of basic amino acids.

Example 18: Combined synthetic approaches

In addition to the above schemes, the coupling approaches described in Examples 8 and 10-17 can be combined, in succession, to synthesize DNA. For example, after a group of codons is prepared as described in Example 8, the scheme described in Example 10 may be used to generate the next set of codons. This process may be continued until DNA of the desired length is achieved.

Alternatively, the trinucleotides generated by any approach may be cleaved from the support using concentrated ammonia at room temperature. The 3'-OH group is then derivatized with allyloxy bis-(diisopropylamino)phosphine to yield the trinucleotide phosphoramidite, and the trinucleotide phosphoramidites are then used as building blocks to synthesize DNA.

Use

The methods of the invention may be used for any application in which nucleic acid synthesis is required. For example, these methods can be used in the synthesis of single-stranded DNA. Frequently, this DNA serve as a template for the synthesis of a complementary DNA strand, which can in turn serves as a template for messenger RNA synthesis.

Because of this application, the methods of the invention find use, for example, in techniques of randomized cassette mutagenesis of proteins, phage display techniques, ribosome display techniques, and protein-nucleic acid fusion techniques.

Codon-randomized DNA can also be used in cellular cultures (*in vivo*) for protein expression, or for *in vitro* applications using, for example, T7 RNA polymerase, and *in vitro* translation systems.

All publications and patents mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent was specifically and individually indicated to be incorporated by reference.

Other Embodiments

From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

What is claimed is:

Claims

1. A method for generating a selected set of codons, said method comprising the steps of:

(a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or mixture thereof, wherein a subset A of said first set is protected with a protecting group A', and a subset B of said first set is protected with a protecting group B', wherein A' and B' are orthogonal protecting groups;

(b) selectively removing said protecting group A' from said subset A;

(c) coupling the products of step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said second set is protected with said protecting group A';

(d) optionally removing said protecting group A' from the products of step (c);

(e) optionally coupling the products of step (d) with a third set of mononucleosides, wherein said third set is protected with said protecting group A';

(f) selectively removing said protecting group B' from said subset B;

(g) coupling the products of step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said fourth set is protected with said protecting group A' or said protecting group B';

(h) optionally selectively removing said protecting group B' from the products of step (g); and

(i) optionally coupling the products of step (h) with a fifth set of mononucleosides, to yield a selected set of codons.

2. The method of claim 1, wherein said selected set of codons comprises at least one codon corresponding to each of the 20 naturally-occurring amino acids.

3. The method of claim 2, wherein each of said codons corresponds to a highly-expressed codon for one of the 20 naturally-occurring amino acids.

4. The method of claim 1, wherein said selected set of codons consists essentially of codons for hydrophobic amino acids, consists essentially of codons for hydrophilic amino acids, consists essentially of codons for basic amino acids, or consists essentially of codons for acidic amino acids.

5. The method of claim 1, wherein fewer than 3%, fewer than 2%, fewer than 1%, fewer than 0.5%, or fewer than 0.1% of said codons correspond to a stop codon.

6. The method of claim 1, wherein steps (a) to (i) take place in the same reaction vessel.

7. The method of claim 1, wherein said protecting groups A' and B' are two different groups and are chosen from an acid-cleavable protecting group, a base-cleavable protecting group, or a fluoride-cleavable protecting group.

8. The method of claim 7, wherein said protecting groups A' and B' are two different groups and are chosen from a dimethoxytrityl group, a fluorenylmethyloxycarbonyl group, or a silyl group.

9. The method of claim 1, wherein each of said codons terminates in a cytidine or a guanosine residue.

10. A method for generating an oligonucleotide from a selected set of codons, said method comprising the steps of:

(a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein a subset A of said first set is protected with a protecting group A', and a subset B of said first set is protected with a protecting group B', wherein A' and B' are orthogonal protecting groups;

(b) selectively removing said protecting group A' from said subset A;

(c) coupling the products of step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said second set is protected with said protecting group A';

(d) optionally removing said protecting group A' from the products of step (c);

(e) optionally coupling the products of step (d) with a third set of mononucleosides, wherein said third set is protected with said protecting group A';

(f) selectively removing said protecting group B' from said subset B;

(g) coupling the products of step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said fourth set is protected with said protecting group A' or said protecting group B';

(h) optionally selectively removing said protecting group B' from the products of step (g);

(i) optionally coupling the products of step (h) with a fifth set of

mononucleosides;

(j) removing the protecting groups from the products of step (g) or (i); and

(k) repeating steps (a) to (j) until an oligonucleotide with the desired length is achieved.

11. The method of claim 10, wherein steps (a) to (k) take place in the same reaction vessel.

12. A method for generating a selected set of codons, said method comprising the steps of:

(a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein a subset A of said first set is protected with a protecting group A', a subset B of said first set is protected with a protecting group B', and a subset C of said first set is protected with a protecting group C', wherein A', B', and C' are orthogonal protecting groups;

(b) selectively removing said protecting group A' from said subset A;

(c) coupling the product formed in step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said second set is protected with said protecting group A';

(d) optionally removing said protecting group A' from the products of step (c);

(e) optionally coupling the products of step (d) with a third set of mononucleosides, wherein said third set of mononucleosides is protected with said protecting group A';

(f) selectively removing said protecting group B' from said subset B;

(g) coupling the products formed in step (f) with a fourth set of

mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said fourth set is protected with said protecting group A' or said protecting group B';

(h) optionally selectively removing said protecting group B' from the products of step (g);

(i) optionally coupling the products of step (h) with a fifth set of mononucleosides, wherein said fifth set is protected with protecting group A';

(j) selectively removing said protecting group C' from said subset C;

(k) coupling the products formed in step (j) with a sixth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein a subset of said sixth set is protected with said protecting group C', and the remainder of said sixth set is protected with protecting group B';

(l) optionally selectively removing said protecting group B' from the products of step (k);

(m) optionally coupling the products of step (l) with a seventh set of mononucleosides, wherein said seventh set is protected with protecting group A' or protecting group B';

(n) selectively removing said protecting group C' from the products of step (m); and

(o) coupling the products of step (n) with an eighth set of mononucleosides, to yield a selected set of codons.

13. The method of claim 12, wherein steps (a) to (o) take place in the same reaction vessel.

14. The method of claim 12, wherein one of said protecting groups A', B', and C' is an acid-cleavable protecting group, one of said protecting groups A', B', and C' is a base-cleavable protecting group, and one of said

protecting groups A', B', and C' is a fluoride-cleavable protecting group.

15. The method of claim 14, wherein one of said protecting groups A', B', and C' is a dimethoxytrityl group, one of said protecting groups A', B', and C' is a fluorenylmethyloxycarbonyl group, and one of said protecting groups A', B', and C' is a silyl group.

16. A method for generating an oligonucleotide from a selected set of codons, said method comprising the steps of:

(a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein a subset A of said first set is protected with a protecting group A', a subset B of said first set is protected with a protecting group B', and a subset C of said first set is protected with a protecting group C', wherein A', B', and C' are orthogonal protecting groups;

(b) selectively removing said protecting group A' from said subset A;

(c) coupling the product formed in step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said second set is protected with said protecting group A';

(d) optionally removing said protecting group A' from the products of step (c);

(e) optionally coupling the products of step (d) with a third set of mononucleosides, wherein said third set is protected with said protecting group A';

(f) selectively removing said protecting group B' from said subset B;

(g) coupling the products formed in step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said fourth set is protected with said protecting group A' or said

protecting group B';

(h) optionally selectively removing said protecting group B' from the products of step (g);

(i) optionally coupling the products of step (h) with a fifth set of mononucleosides, wherein said fifth set is protected with protecting group A';

(j) selectively removing said protecting group C' from said subset C;

(k) coupling the products formed in step (j) with a sixth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein a subset of said sixth set is protected with said protecting group C', and the remainder of said sixth set is protected with protecting group B';

(l) optionally selectively removing said protecting group B' from the products of step (k);

(m) optionally coupling the products of step (l) with a seventh set of mononucleosides, wherein said seventh set is protected with protecting group A' or protecting group B';

(n) selectively removing said protecting group C' from the products of step (m);

(o) coupling the products of step (n) with an eighth set of mononucleosides;

(p) removing the protecting groups from the products of step (o); and

(q) repeating steps (a) to (p) until an oligonucleotide with the desired length is achieved.

17. The method of claim 16, wherein steps (a) to (q) take place in the same reaction vessel.

18. A method for generating, in the same reaction vessel, a selected set of codons, said method comprising the steps of:

(a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof;

(b) adding a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof;

(c) optionally adding a third set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof; and

(d) optionally repeating step (c) to yield a selected set of codons, wherein said selected set includes at least one codon having A or G at the third codon position, and wherein fewer than 3% of the codons in said selected set correspond to a stop codon.

19. The method of claim 18, wherein said selected set of codons includes at least one codon for each of the 20 naturally-occurring amino acids.

20. The method of claim 19, wherein each of said codons corresponds to a highly-expressed codon for one of the 20 naturally-occurring amino acids.

21. The method of claim 18, wherein said selected set of codons consists essentially of codons for basic amino acids or consists essentially of codons for hydrophobic amino acids.

22. The method of claim 18, wherein fewer than 2%, fewer than 1%, fewer than 0.5%, or fewer than 0.1% of said codons correspond to a stop codon.

23. The method of claim 18, wherein each of said codons terminates in a cytidine or a guanosine residue.

24. A method for generating an oligonucleotide from a selected set of codons, said method comprising the steps of:

(a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or mixture thereof;

(b) adding a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof;

(c) optionally adding a third set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof;

(d) optionally repeating step (c) to yield a selected set of codons, wherein said selected set includes at least one codon having A or G at the third codon position, wherein fewer than 3% of the codons in said selected set correspond to a stop codon, and wherein steps (a), (b), (c) and (d) occur in the same reaction vessel; and

(e) repeating steps (a) to (d) until an oligonucleotide of the desired length is achieved.

25. The method of claim 24, wherein said selected set of codons includes at least one codon for each of the 20 naturally-occurring amino acids.

26. The method of claim 24, wherein fewer than 2% of said codons correspond to a stop codon.

Figure 1

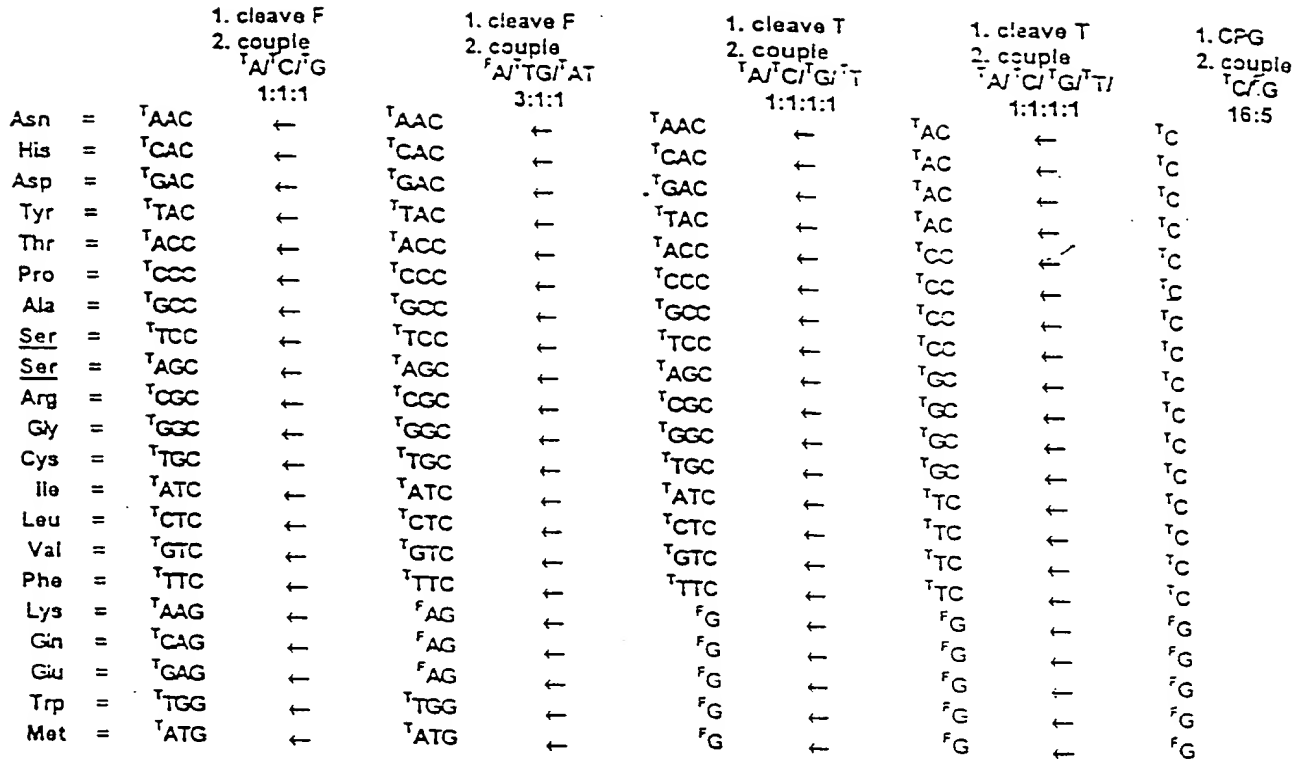
1. cleave F
2. couple
TA/TC/TG
1:1:1
1. cleave F
2. couple
FA/TTG/TAT
3:1:1
1. cleave T
2. couple
TA/TC/TG/TT
1:1:1:1
1. cleave T
2. couple
TA/TC/TG/TT/
1:1:1:1
1. CPG
2. couple
TC/FG
16:5
Asn = TAAC
His = TCAC
Asp = TGAC
Tyr = TTAC
Thr = TACC
Pro = TCCC
Ala = TGCC
Ser = TTCC
Ser = TAGC
Arg = TCGC
Gly = TGGC
Cys = TTGC
Ile = TATC
Leu = TCTC
Val = TGTC
Phe = TTTC
Lys = TAAG
Gln = TCAG
Glu = TGAG
Trp = TTGG
Met = TATG
TAAC
TCAC
TGAC
TTAC
TACC
TCCC
TGCC
TTCC
TAGC
TCGC
TGGC
TTGC
TATC
TCTC
TGTC
TTTC
FAG
FAG
FAG
TTGG
TATG
TAAC
TCAC
TGAC
TTAC
TACC
TCCC
TGCC
TTCC
TAGC
TCGC
TGGC
TTGC
TATC
TCTC
TGTC
TTTC
FG
FG
FG
FG
FG
TAC
TAC
TAC
TAC
TCC
TCC
TCC
TCC
TGC
TGC
TGC
TGC
TTC
TTC
TTC
TTC
FG
FG
FG
FG
FG
TC
TC
TC
TC
TC
TC
TC
TC
TC
TC
TC
TC
TC
TC
TC
TC
FG
FG
FG
FG
FG

Figure 2

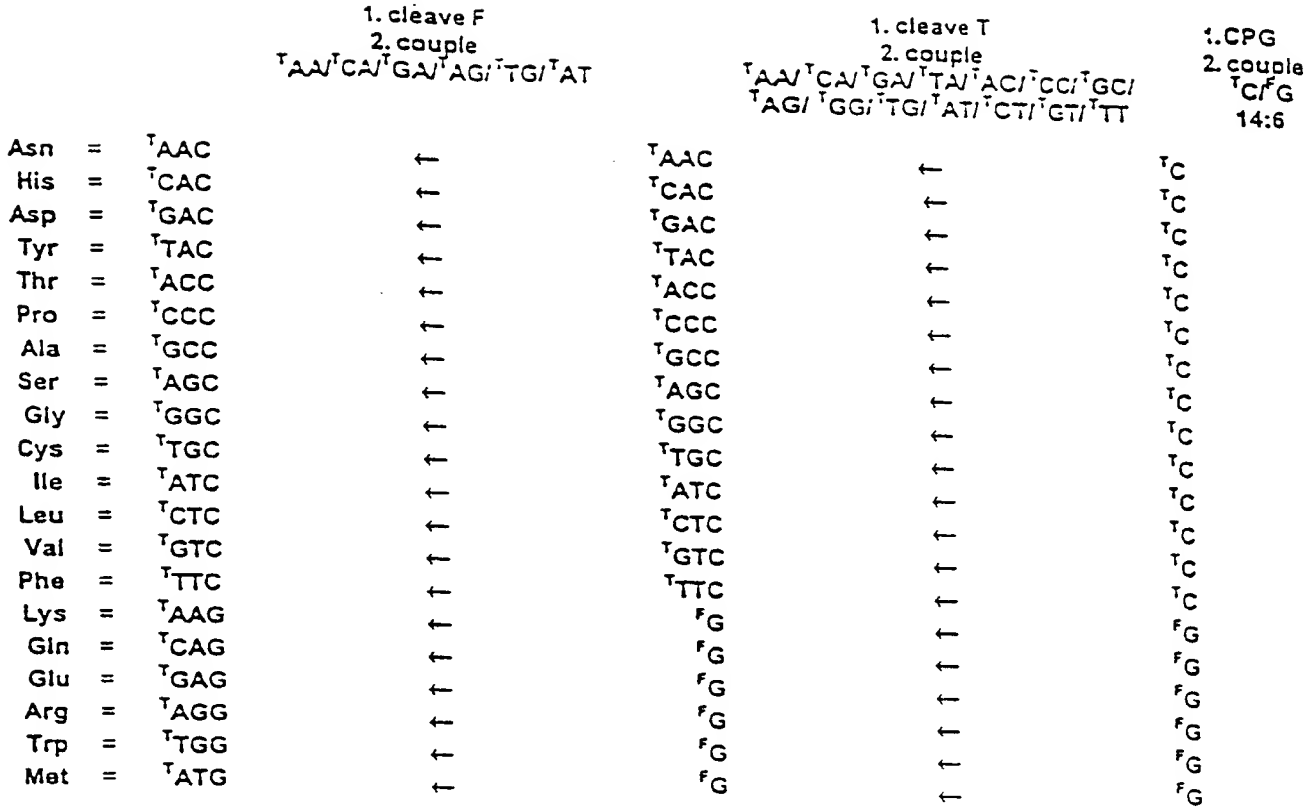
1. cleave F
2. couple
TAA/TCA/TGA/TAG/TTG/TAT
1. cleave T
2. couple
TAA/TCA/TGA/TTA/TAC/TCC/TGC/
TAG/TGG/TTG/TAT/TCT/TGT/TTT
1. CPG
2. couple
TC/FG
14:6
Asn = TAAC
His = TCAC
Asp = TGAC
Tyr = TTAC
Thr = TACC
Pro = TCCC
Ala = TGCC
Ser = TAGC
Gly = TGGC
Cys = TTGC
Ile = TATC
Leu = TCTC
Val = TGTC
Phe = TTTC
Lys = TAAG
Gln = TCAG
Glu = TGAG
Arg = TAGG
Trp = TTGG
Met = TATG
TAAC
TCAC
TGAC
TTAC
TACC
TCCC
TGCC
TAGC
TGGC
TTGC
TATC
TCTC
TGTC
TTTC
FG
FG
FG
FG
FG
FG
TC
TC
TC
TC
TC
TC
TC
TC
TC
TC
TC
TC
FG
FG
FG
FG
FG
FG

Figure 3

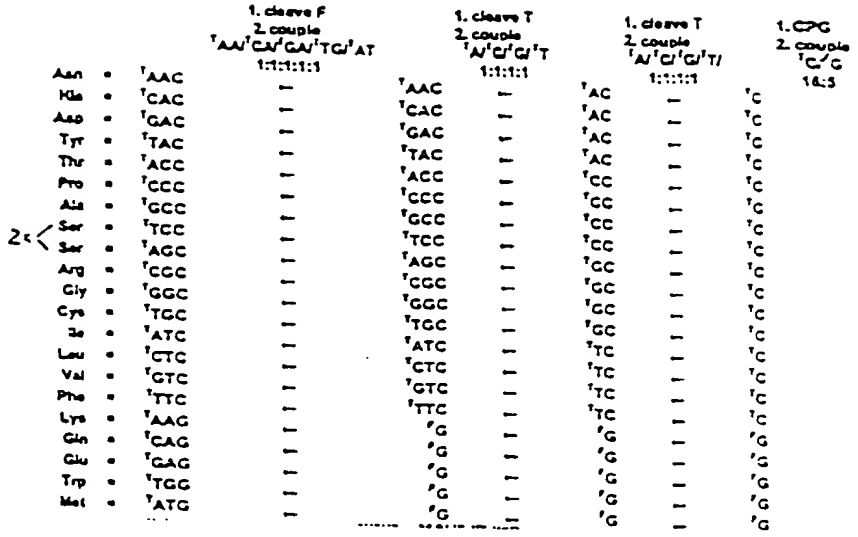

Figure 4

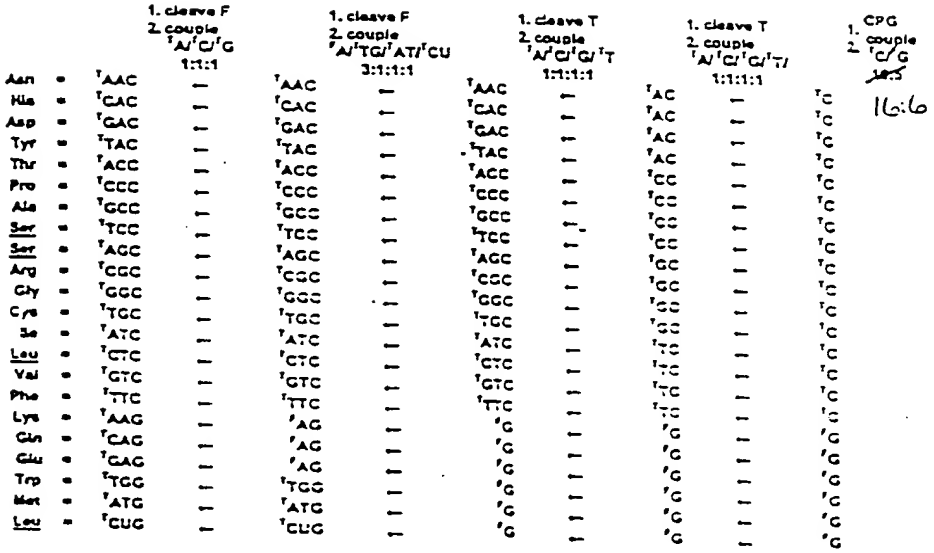

Figure 5

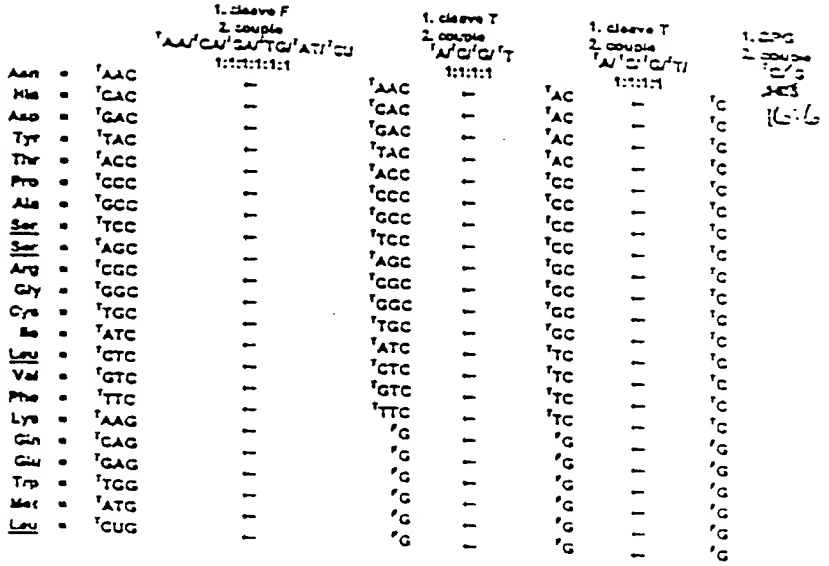

Figure 6

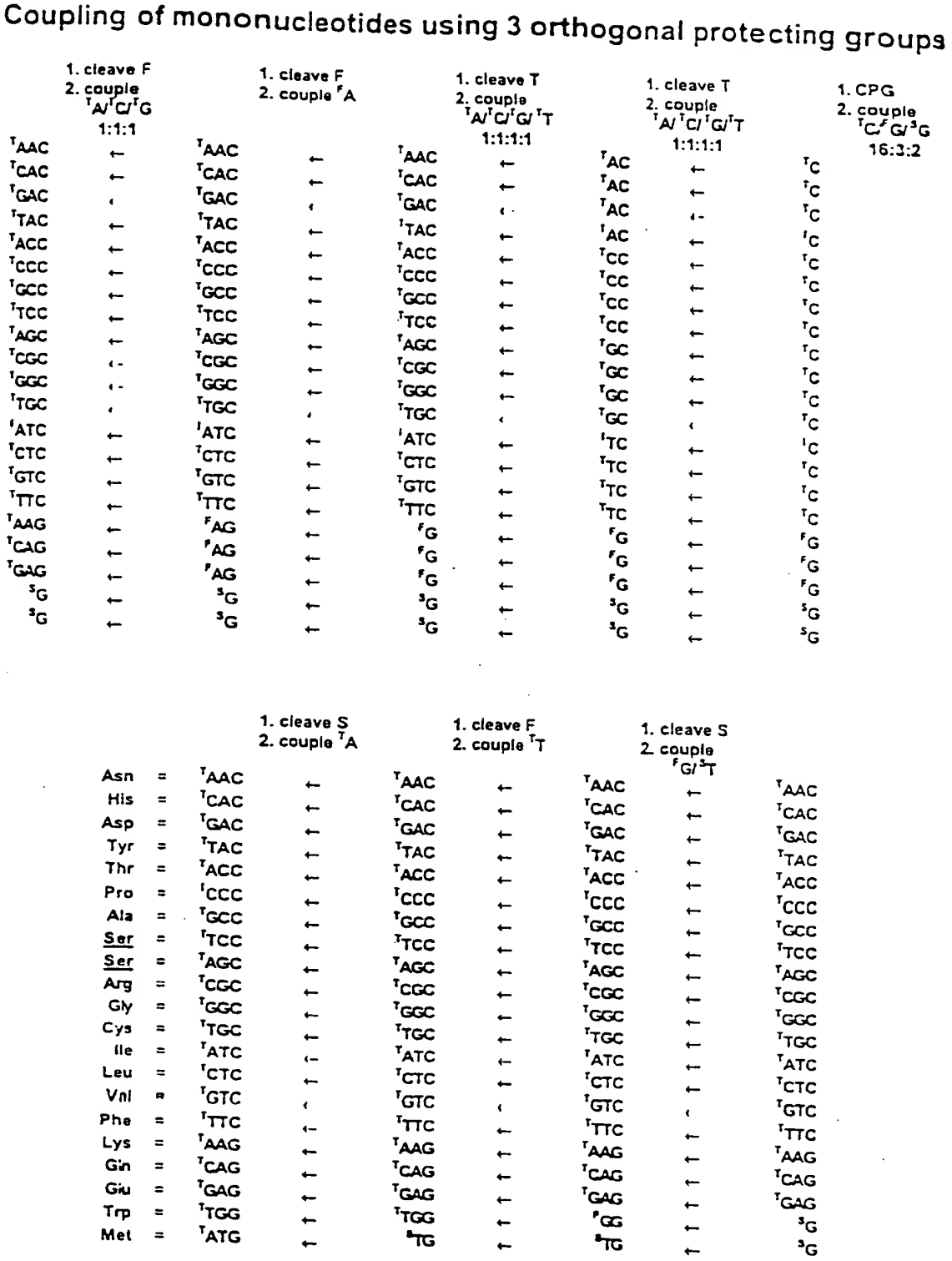
Coupling of mononucleotides using 3 orthogonal protecting groups
1. cleave F
2. couple TA/TC/TG 1:1:1
1. cleave F
2. couple FA
1. cleave T
2. couple TA/TC/TG/TT 1:1:1:1
1. cleave T
2. couple TA/TC/TG/TT 1:1:1:1
1. CPG
2. couple TC/FG/3G 16:3:2
1. cleave S
2. couple TA
1. cleave F
2. couple TT
1. cleave S
2. couple FG/3T
Asn = TAAC
His = TCAC
Asp = TGAC
Tyr = TTAC
Thr = TACC
Pro = TCCC
Ala = TGCC
Ser = TTCC
Ser = TAGC
Arg = TCGC
Gly = TGGC
Cys = TTGC
Ile = TATC
Leu = TCTC
Val = TGTC
Phe = TTTC
Lys = TAAG
Gln = TCAG
Glu = TGAG
Trp = TTGG
Met = TATG

Figure 7

Synthesis of codon-randomized DNA coding for hydrophobic amino acids only

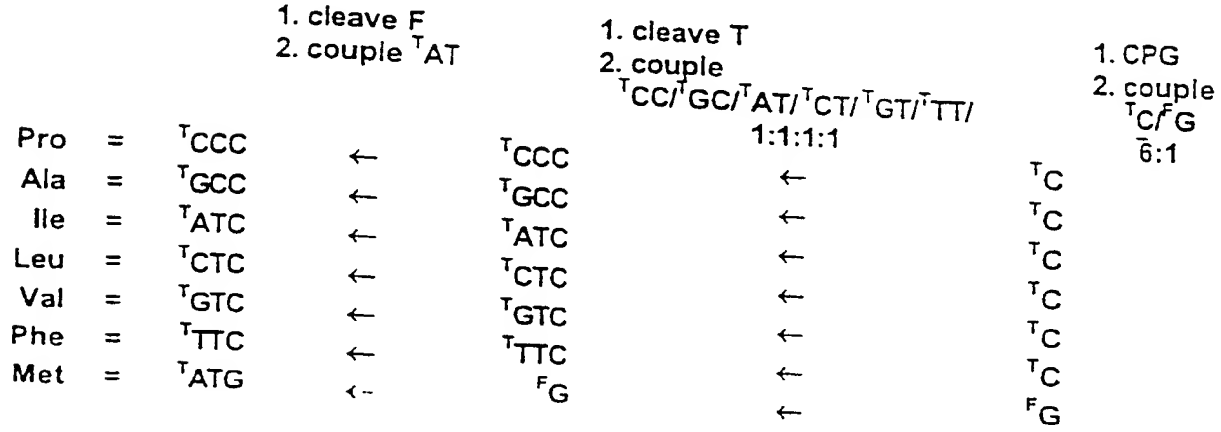

Figure 8

Synthesis of codon-randomized DNA biased towards hydrophobic amino acid codons (Ile, Leu, Val, Phe)

			1. cleave F 2. couple $^{T}AA/^{T}CA/^{T}GA/^{T}TG/^{T}AT$ 1:1:1:1:1	1. cleave T 2. couple $^{T}A/^{T}C/^{T}G/^{T}T$ 1:1:1:1	1. cleave T 2. couple $^{T}A/^{T}C/^{T}G/^{T}T/$ 1:1:1:2	1. CPG 2. couple $^{T}C/^{F}G$ 16:5
Asn	=	^{T}AAC	←	^{T}AAC ←	^{T}AC ←	^{T}C
His	=	^{T}CAC	←	^{T}CAC ←	^{T}AC ←	^{T}C
Asp	=	^{T}GAC	←	^{T}GAC ←	^{T}AC ←	^{T}C
Tyr	=	^{T}TAC	←	^{T}TAC ←	^{T}AC ←	^{T}C
Thr	=	^{T}ACC	←	^{T}ACC ←	^{T}CC ←	^{T}C
Pro	=	^{T}CCC	←	^{T}CCC ←	^{T}CC ←	^{T}C
Ala	=	^{T}GCC	←	^{T}GCC ←	^{T}CC ←	^{T}C
Ser	=	^{T}TCC	←	^{T}TCC ←	^{T}CC ←	^{T}C
Ser	=	^{T}AGC	←	^{T}AGC ←	^{T}GC ←	^{T}C
Arg	=	^{T}CGC	←	^{T}CGC ←	^{T}GC ←	^{T}C
Gly	=	^{T}GGC	←	^{T}GGC ←	^{T}GC ←	^{T}C
Cys	=	^{T}TGC	←	^{T}TGC ←	^{T}GC ←	^{T}C
2 Ile	=	$2\ ^{T}ATC$	←	$2\ ^{T}ATC$ ←	$2\ ^{T}TC$ ←	^{T}C
2 Leu	=	$2\ ^{T}CTC$	←	$2\ ^{T}CTC$ ←	$2\ ^{T}TC$ ←	^{T}C
2 Val	=	$2\ ^{T}GTC$	←	$2\ ^{T}GTC$ ←	$2\ ^{T}TC$ ←	^{T}C
2 Phe	=	$2\ ^{T}TTC$	←	$2\ ^{T}TTC$ ←	$2\ ^{T}TC$ ←	^{T}C
Lys	=	^{T}AAG	←	^{F}G ←	^{F}G ←	^{F}G
Gln	=	^{T}CAG	←	^{F}G ←	^{F}G ←	^{F}G
Glu	=	^{T}GAG	←	^{F}G ←	^{F}G ←	^{F}G
Trp	=	^{T}TGG	←	^{F}G ←	^{F}G ←	^{F}G
Met	=	^{T}ATG	←	^{F}G ←	^{F}G ←	^{F}G

Figure 9

Synthesis of codon-randomized DNA biased towards basic amino acid codons (His, Lys, Arg)

			1. cleave F 2. couple $^{T}AA/^{T}CA/^{T}GA/^{T}AG/^{T}TG/^{T}AT$ 2:1:1:2:1:1		1. cleave T 2. couple $^{T}AA/^{T}CA/^{T}GA/^{T}TA/^{T}AC/^{T}CC/^{T}GC/$ $^{T}AG/^{T}GG/^{T}TG/^{T}AT/^{T}CT/^{T}GT/^{T}TT$ 1:2:1:1:1:1:1:1:1:1:1:1:1:1:1		1. CPG 2. couple $^{T}C/^{T}G$ 14:6
Asn	=	^{T}AAC	←	^{T}AAC	←	^{T}C	
2 His	=	$2\ ^{T}CAC$	←	$2\ ^{T}CAC$	←	^{T}C	
Asp	=	^{T}GAC	←	^{T}GAC	←	^{T}C	
Tyr	=	^{T}TAC	←	^{T}TAC	←	^{T}C	
Thr	=	^{T}ACC	←	^{T}ACC	←	^{T}C	
Pro	=	^{T}CCC	←	^{T}CCC	←	^{T}C	
Ala	=	^{T}GCC	←	^{T}GCC	←	^{T}C	
Ser	=	^{T}AGC	←	^{T}AGC	←	^{T}C	
Gly	=	^{T}GGC	←	^{T}GGC	←	^{T}C	
Cys	=	^{T}TGC	←	^{T}TGC	←	^{T}C	
Ile	=	^{T}ATC	←	^{T}ATC	←	^{T}C	
Leu	=	^{T}CTC	←	^{T}CTC	←	^{T}C	
Val	=	^{T}GTC	←	^{T}GTC	←	^{T}C	
Phe	=	^{T}TTC	←	^{T}TTC	←	^{T}C	
2 Lys	=	$2\ ^{T}AAG$	←	^{F}G	←	^{F}G	
Gln	=	^{T}CAG	←	^{F}G	←	^{F}G	
Glu	=	^{T}GAG	←	^{F}G	←	^{F}G	
2 Arg	=	$2\ ^{T}AGG$	←	^{F}G	←	^{F}G	
Trp	=	^{T}TGG	←	^{F}G	←	^{F}G	
Met	=	^{T}ATG	←	^{F}G	←	^{F}G	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/22436

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 1/00, 1/02; C07C 255/11, 255/49

US CL : 536/25.3, 25.34, 26.1; 558/423, 435; 568/597, 598

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/25.3, 25.34, 26.1; 558/423, 435; 568/597, 598

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,E	US 5,889,136 A (SCARINGE et al.) 30 March 1999, col. 2-6.	1-26
Y	US 5,703,218 A (URDEA et al.) 30 December 1997, col. 3-6.	1,12,16

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 03 JANUARY 2000	Date of mailing of the international search report 02 FEB 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer HOWARD OWENS Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)★

VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGA. _ATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C07H 1/00, 1/02, C07C 255/11, 255/49

A1

(11) International Publication Number: WO 00/18778

(43) International Publication Date: 6 April 2000 (06.04.00)

(21) International Application Number: PCT/US99/22436

(22) International Filing Date: 28 September 1999 (28.09.99)

(30) Priority Data:
60/102,299 29 September 1998 (29.09.98) US

(71) Applicant: PHYLOS, INC. [US/US]; 128 Spring Street, Lexington, MA 02421 (US).

(72) Inventors: LOHSE, Peter; 50 Golden Ball Road, Weston, MA 02493 (US). KUIMELIS, Robert, G.; 21 Malbert Road, Brighton, MA 02135 (US).

(74) Agent: ELBING, Karen, L.; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110–2214 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SYNTHESIS OF CODON RANDOMIZED NUCLEIC ACIDS

(57) Abstract

A method for generating a selected set of codons is disclosed; the method includes the steps of: (a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or mixture thereof, where a subset A of the first set is protected with a protecting group A', and a subset B of the first set is protected with a protecting group B', where A' and B' are orthogonal protecting groups; (b) selectively removing the protecting group A' from subset A; (c) coupling the products of step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the second set is protected with protecting group A'; (d) optionally removing protecting group A' from the products of step (c); (e) optionally coupling the products of step (d) with a third set of mononucleosides, where the third set is protected with protecting group A'; (f) selectively removing the protecting group B' from subset B; (g) coupling the products of step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the fourth set is protected with protecting group A' or protecting group B'; (h) optionally selectively removing protecting group B' from the products of step (g); and (i) optionally coupling the products of step (h) with a fifth set of mononucleosides, to yield a selected set of codons.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

SYNTHESIS OF CODON RANDOMIZED NUCLEIC ACIDS

Background of the Invention

The invention relates to methods for chemically synthesizing DNA or RNA.

Pharmaceutical research relies, in part, on the identification of novel proteins with desired functions and properties. To identify proteins or peptides with improved properties, derivatives of known proteins and peptides can be prepared using methods such as oligonucleotide-directed mutagenesis. Proteins with desired functions can also be selected from pools of randomly synthesized proteins, including proteins which are generated from random DNA template libraries.

DNA libraries, in turn, may also be generated using a variety of techniques. Such DNA libraries can be synthesized on a solid support (e.g., a CPG support), in a liquid phase, or in a combination solid-liquid phase (e.g., a PEG support). Most commonly, DNA libraries are prepared using a standard DNA synthesizer and a random mixture of all 4 nucleotides in each coupling step. By this approach, the trinucleotides, or codons, that correspond to the different amino acids, are randomly generated. This codon randomized DNA can then be transcribed into RNA, which is in turn used to synthesize polypeptides; the approach described above thus provides a means for generating a wide variety of DNA sequences and proteins products.

Although it is commonly utilized, the random generation of DNA by conventional techniques can have disadvantages. For example, methods that rely on completely random generation of codons generally suffer from limited control over the synthesis of polypeptides generated from this DNA. The presence of weakly expressed codons in the random product mixture, for

example, lowers the efficiency with which the DNA is translated. Furthermore, a small subset of randomly generated codons (approximately 3 out of 64) corresponds to a stop codon. As the presence of stop codons terminates protein synthesis, protein libraries generated from randomly generated DNA templates can sometimes exhibit low yields of full-length proteins. In addition, methods that rely on the completely random generation of DNA do not allow for a bias for a selected group of amino acids, for example, hydrophobic amino acids.

Summary of the Invention

In one aspect, the invention features a method for generating a selected set of codons; the method includes the steps of: (a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where a subset A of the first set is protected with a protecting group A', and a subset B of the first set is protected with a protecting group B', where A' and B' are orthogonal protecting groups; (b) selectively removing protecting group A' from subset A; (c) coupling the products of step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the second set is protected with protecting group A'; (d) optionally removing protecting group A' from the products of step (c); (e) optionally coupling the products of step (d) with a third set of mononucleosides, where the third set is protected with protecting group A'; (f) selectively removing protecting group B' from subset B; (g) coupling the products of step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the fourth set is protected with protecting group A' or protecting group B'; (h) optionally selectively removing protecting group B' from the products of step (g); and (i) optionally coupling the products of step (h) with a fifth set of mononucleosides, to yield a selected set of codons.

In preferred methods, the selected set includes at least one codon

corresponding to each of the 20 naturally-occurring amino acids; preferably, each of these codons corresponds to a highly expressed codon for one of the naturally-occurring amino acids. The selected set may also consist of trinucleotides coding only for a class of amino acids, e.g., hydrophobic amino acids, hydrophilic amino acids, basic amino acids, or acidic amino acids. In another preferred method, the selected set may consist of trinucleotides coding for a mixture of amino acids, e.g., acidic and basic amino acids.

Preferably, fewer than 3% of the codons correspond to a stop codon; more preferably, fewer than 2%, 1%, 0.5% or 0.1%, of the codons correspond to a stop codon. In preferred methods, steps (a) to (i) take place in the same reaction vessel; in addition, protecting groups A' and B' are two different groups and are preferably chosen from an acid-cleavable protecting group (for example a dimethoxytrityl group), a base-cleavable protecting group (for example, a fluorenylmethyloxycarbonyl group), or a fluoride-cleavable protecting group (for example, a silyl group). In other preferred methods, each of the codons terminates in a cytidine or a guanosine residue.

In a second aspect, the invention features a method for generating an oligonucleotide from a selected set of codons; the method includes the steps of: (a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where a subset A of the first set is protected with a protecting group A', and a subset B of the first set is protected with a protecting group B', where A' and B' are orthogonal protecting groups; (b) selectively removing protecting group A' from subset A; (c) coupling the products of step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the second set is protected with protecting group A'; (d) optionally removing protecting group A' from the products of step (c); (e) optionally coupling the products of step (d) with a third set of mononucleosides, where the third set is protected with protecting group A'; (f)

selectively removing protecting group B' from subset B; (g) coupling the products of step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the fourth set is protected with protecting group A' or protecting group B'; (h) optionally selectively removing protecting group B' from the products of step (g); (i) optionally coupling the products of step (h) with a fifth set of mononucleosides; (j) removing the protecting groups from the products of step (g) or (i); and (k) repeating steps (a) to (j) until an oligonucleotide with the desired length is achieved. Preferably, steps (a) to (k) take place in the same reaction vessel.

In a third aspect, the invention features a method for generating a selected set of codons including the steps of: (a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where a subset A of the first set is protected with a protecting group A', a subset B of the first set is protected with a protecting group B', and a subset C of the first set is protected with a protecting group C', where A', B', and C' are orthogonal protecting groups; (b) selectively removing the protecting group A' from the subset A;(c) coupling the product formed in step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the second set is protected with the protecting group A'; (d) optionally removing the protecting group A' from the products of step (c); (e) optionally coupling the products of step (d) with a third set of mononucleosides, where the third set of mononucleosides is protected with the protecting group A'; (f) selectively removing the protecting group B' from the subset B; (g) coupling the products formed in step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the fourth set is protected with the protecting group A' or the protecting group B'; (h) optionally selectively removing the protecting group B' from the products of step (g); (i) optionally coupling the products of step (h) with a fifth set of

mononucleosides, where the fifth set is protected with protecting group A'; (j) selectively removing the protecting group C' from the subset C; (k) coupling the products formed in step (j) with a sixth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where a subset of the sixth set is protected with the protecting group C', and the remainder of the sixth set is protected with protecting group B'; (l) optionally selectively removing the protecting group B' from the products of step (k); (m) optionally coupling the products of step (l) with a seventh set of mononucleosides, where the seventh set is protected with protecting group A' or protecting group B'; (n) selectively removing the protecting group C' from the products of step (m); and (o) coupling the products of step (n) with an eighth set of mononucleosides, to yield a selected set of codons.

In preferred methods, steps (a) to (o) take place in the same reaction vessel. In addition, one of the protecting groups A', B', and C' is preferably an acid-cleavable protecting group (for example a dimethoxytrityl group), another of the protecting groups A', B', and C' is preferably a base-cleavable protecting group (for example, a fluorenylmethyloxycarbonyl group), and the last of the protecting groups A', B', and C' is preferably a fluoride-cleavable protecting group (for example, a silyl group).

In a fourth aspect, the invention features a method for generating an oligonucleotide from a selected set of codons including the steps of: (a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where a subset A of the first set is protected with a protecting group A', a subset B of the first set is protected with a protecting group B', and a subset C of the first set is protected with a protecting group C', where A', B', and C' are orthogonal protecting groups; (b) selectively removing the protecting group A' from the subset A; (c) coupling the product formed in step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture

thereof, where the second set is protected with the protecting group A'; (d) optionally removing the protecting group A' from the products of step (c); (e) optionally coupling the products of step (d) with a third set of mononucleosides, where the third set is protected with the protecting group A'; (f) selectively removing the protecting group B' from the subset B; (g) coupling the products formed in step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where the fourth set is protected with the protecting group A' or the protecting group B'; (h) optionally selectively removing the protecting group B' from the products of step (g); (i) optionally coupling the products of step (h) with a fifth set of mononucleosides, where the fifth set is protected with protecting group A'; (j) selectively removing the protecting group C' from the subset C; (k) coupling the products formed in step (j) with a sixth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, where a subset of the sixth set is protected with the protecting group C', and the remainder of the sixth set is protected with protecting group B'; (l) optionally selectively removing the protecting group B' from the products of step (k); (m) optionally coupling the products of step (l) with a seventh set of mononucleosides, where the seventh set is protected with protecting group A' or protecting group B'; (n) selectively removing the protecting group C' from the products of step (m); (o) coupling the products of step (n) with an eighth set of mononucleosides; (p) removing the protecting groups from the products of step (o); and (q) repeating steps (a) to (p) until an oligonucleotide with the desired length is achieved. Preferably, steps (a) to (q) take place in the same reaction vessel.

In a fifth aspect, the invention features a method for generating, in the same reaction vessel, a selected set of codons; the method includes the steps of: (a) providing a first set of mononucleosides, mononucleotides, or dinucleotides, or mixture thereof; (b) adding a second set of mononucleosides,

mononucleotides, dinucleotides, or a mixture thereof; (c) optionally adding a third set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof; and (d) optionally repeating step (c) to yield a selected set of codons. The selected set includes at least one codon having A or G at the third codon position; fewer than 3% of the codons in the selected set correspond to a stop codon.

In preferred methods, the selected set includes at least one codon for each of the 20 naturally-occurring amino acids; preferably, each codon corresponds to a highly-expressed codon for one of the naturally-occurring amino acids. In other preferred methods, the selected set may consist of one class of codons, e.g., hydrophobic amino acids. In another preferred method, the selected set may consist of trinucleotides coding for a mixture of amino acids, e.g., acidic and basic amino acids. Preferably, fewer than 2% of the codons correspond to a stop codon; more preferably, fewer than 1%, 0.5%, or 0.1%, of the codons correspond to a stop codon. In still other preferred methods, each of the codons terminates in a cytidine or a guanosine residue.

Any combination of couplings of mononucleosides, mononucleotides, and dinucleotides may be used to generated codons, which are trinucleotides. For example, dinucleotides may be coupled with mononucleosides. Dinucleotides would not be coupled with dinucleotides, as that would generated tetranucleotides.

In a sixth aspect, the invention features a method for generating an oligonucleotide from a selected set of codons. The method includes the steps of: (a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof; (b) adding a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof; (c) optionally adding a third set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof; and (d) optionally repeating step (c) to yield

a selected set of codons that includes at least one codon having A or G at the third codon position and in which fewer than 3% of the codons correspond to a stop codon. Steps (a), (b), (c), and (d) occur in the same reaction vessel; these steps are repeated until an oligonucleotide of the desired length is achieved. Preferably, the selected set includes at least one codon for each of the 20 naturally-occurring amino acids, and fewer than 2% of the codons correspond to a stop codon.

By "nucleoside" is meant any sugar-base moiety, including sugar-base moieties in which one or more nitrogen atoms of the nitrogenous bases are protected, and/or in which the 5'-OH of the sugar is protected. "Nucleosides" also include nucleoside phosphoramidites and protected phosphoramidites.

By "nucleotide" is meant any sugar-phosphate-base moiety, as well as any derivatized sugar-phosphate-base moiety. One or more nitrogen atoms of the nitrogenous bases can be protected, and/or the 5'-OH of the sugar can be protected. Dinucleotides can include dinucleotide phosphoramidites; in addition, the internucleotide linkage may be protected.

By "oligonucleotide" is meant either a DNA sequence or an RNA sequence; by "nucleic acid" is meant either DNA or RNA.

By "highly-expressed codons" are meant the codons present in higher than normal abundance in highly expressed genes.

By "stop codon" is meant one of the DNA codons TAA, TGA, and TAG; and the RNA codons UAA, UGA, and UAG.

By a "selected set of codons" is meant a set of trinucleotide sequences where each trinucleotide has an assigned representation in the set. For example, a selected set of codons may be a set that contains at least one codon for each of the naturally occuring amino acids (e.g., AAC : CAC : GAC : TAC: ACC : CCC : GCC : TCC : AGC : CGC : GGC : TGC : ATC : CTC : GTC : TTC : AAG : CAG : GAG : TGG : ATG =

1:1:1:1:1:1:1:1:1:1:1:1:1:1:1:1:1:1:1:1:1). Alternatively, a selected set of codons may be a set that contains at least one codon for each of the naturally occurring amino acids, and in which some hydrophobic amino acids (e.g., Val, Leu, Ile, Phe) are twice as abundant (e.g., AAC : CAC : GAC : TAC : ACC : CCC : GCC : TCC : AGC : CGC : GGC :TGC : ATC : CTC : GTC : TTC : AAG : CAG : GAG : TGG : ATG = 1:1:1:1:1:1:1:1:1:1:2:2:2:2:2:1:1:1:1:1). A selected set of codons may also be a set that is biased towards basic amino acids (e.g., His, Lys, Arg). The composition of such a set can be, e.g., AAC : CAC : GAC : TAC : ACC : CCC : GCC : AGC : GGC : TGC : ATC : CTC : GTC : TTC : AAG : CAG : GAG : AGG : TGG : ATG = 1:2:1:1:1:1:1:1:1:1:1:1:2:1:1:2:1:1. Alternatively, a selected set of codons may be a set in which all of the codons code for hydrophobic amino acids (e.g, Pro, Ala, Ile, Leu, Val, Phe, Met) in equal distribution (e.g., CCC : GCC : ATC : CTC : GTC : TTC : ATG = 1:1:1:1:1:1:1).

The nucleosides and nucleotides used herein are often referred to with shorthand designations, in which the protecting group of the 5'-OH is superscripted. For example, "^{T}C" is used to represent N^4-benzoyl 5'-*O*-(4,4'-dimethoxytrityl) 2'-deoxycytidine or N^4-benzoyl 5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxycytidine, and "^{F}G" is used to represent N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl] 2'-deoxyguanosine or N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl] 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyguanosine.

The present invention provides a number of advantages over conventional techniques of nucleic acid synthesis. For example, the methods described herein provide control over codon format (trinucleotide sequences) as well as control over the representation of the codon in a selected set. The invention can therefore generate sets of codons that contain at least one codon for each of the naturally-occurring amino acids and, importantly, can generate

libraries that are substantially free of stop codons. The invention also provides DNA consisting essentially of highly-expressed codons and, as noted above, free of stop codons, which can be efficiently translated to polypeptides of desired lengths. Furthermore, control over codon representation allows for the synthesis of DNA templates that can be used to generate proteins rich in selected amino acids, for example, hydrophobic amino acids, which can be instrumental in protein design techniques.

Brief Description of the Drawings

FIGURES 1, 2, 3, 4, 5, 6, 7, 8, and 9 are each illustrations of coupling sequences for the synthesis of codon libraries.

Description of the Preferred Embodiments

The sequence of bases in DNA and its RNA counterpart determines the sequence of the amino acids in the protein synthesized from this DNA. Sequences of three bases, referred to as codons, correspond to different amino acids. During translation, these codons are read from the 5' end to the 3' end; the resulting protein has an amino acid sequence that corresponds to the sequence of codons.

Three DNA codons, TAA, TAG, and TGA (which correspond to the RNA codons UAA, UAG, and UGA) do not code for any amino acids. Instead, these codons signal release factors to terminate protein synthesis. The presence of stop codons therefore leads to termination of protein synthesis before the entire DNA sequence is translated.

The invention features convenient methods for the controlled synthesis of codon randomized nucleic acids, such as DNA, in which the presence of stop codons can be avoided. If desired, the DNA strand can be used as a template for the synthesis of a complementary DNA strand, which in

turn can serve as a template for the synthesis of the corresponding messenger RNA. Alternatively, messenger RNA can be synthesized directly using the methods of the invention.

As described in more detail below, in the present approach, a desired set of codons, as well as the desired frequency of each codon in the set, is first chosen. The set can include, for example, the most highly-expressed codons for each of the 20 naturally-occurring amino acids, in equal distribution. Highly expressed DNA codons in eukaryotic translation systems typically exhibit either 2'-deoxycytidine (C) or 2'-deoxyguanosine (G) at the 3' end (that is, at the third codon position). Another desired set can include, for example, at least one codon for each of the 20 naturally-occurring amino acids, and in which hydrophobic amino acids are twice as abundant.

In addition, it is generally desirable to omit, from the codon mixture, the known stop codons, TAA, TGA, and TAG, for DNA synthesis, or UAA, UGA, and UAG for RNA synthesis. The omission of these codons from the synthetic library maximizes the likelihood that protein translation is not aborted and that proteins of desired length are generated.

In one particular example, a selected set of highly expressed codons for all 20 naturally-occurring amino acids can be prepared in which all of the codons have a C or G at the third position. Two of the three stop codons, TAA and TGA, are therefore readily excluded from this set.

In addition, none of the stop codons has a C at the third position; codons ending in C can therefore be randomly generated without the introduction of stop codons. The generation of a set of codons ending in C can produce codons for fifteen of the naturally-occurring amino acids.

The exclusion of the stop codon TAG (or, for RNA synthesis, UAG) is more complicated because the best expressed codons for the amino acids Lys, Gln, Glu, Trp, and Met each have a G at position three. The simultaneous

generation of codons for these amino acids, and the exclusion of the stop codon TAG, requires a strategic coupling sequence. Several such coupling sequences are described in detail below as part of the present invention.

The present invention also features methods for generating libraries of codons by using nucleosides and nucleotides with different 5'-protecting groups as building blocks. In preferred methods, the different 5'-protecting groups can be cleaved under orthogonal conditions. In other words, the conditions for cleaving one 5'-protecting group do not cleave the other 5'-protecting groups. An example of one pair of orthogonal protecting groups includes a dimethoxytrityl group (DMT or T), which is cleaved under acidic conditions, and a fluorenylmethyloxycarbonyl group (Fmoc or F), which is cleaved under basic conditions. Another example of a set of orthogonal protecting groups is the set including a dimethoxytrityl group (DMT or T), which is cleaved under acidic conditions, a fluorenylmethyloxycarbonyl group (Fmoc or F), which is cleaved under basic conditions, and a silyl group (S), which is cleaved with fluoride.

In one particular method, a mixture of nucleosides, some of which are protected with a DMT group, and some of which are protected with a Fmoc group, is treated with acid. The DMT-protected nucleosides are deprotected, while the Fmoc protected nucleosides remain protected. When nucleotides are added to this mixture, they couple only with the deprotected nucleosides, allowing for coupling specificity.

In one example, a mixture of N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxycytidine (^{T}C) and N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl]-2'-deoxyguanosine (^{F}G) is treated with acid. The DMT group is cleaved from the C mononucleosides, thus leaving them free to couple with nucleotides. The Fmoc of the G mononucleosides remains attached. Since none of the stop codons end in C, trinucleotides may be randomly generated at this step without

the introduction of stop codons. By this technique, successive coupling steps produce a mixture of trityl-protected trinucleotides ending in C. Throughout the coupling steps, the G mononucleosides remain protected and therefore unreactive.

Codons with G at the third position are then prepared. The Fmoc groups are cleaved with base. The coupling sequence is designed to avoid synthesis of the codon TAG (or UAG).

In yet further preferred embodiments, similar processes are carried out using silyl protecting groups.

Examples of several different coupling schemes are given below. It is to be understood, however, that the invention encompasses additional coupling schemes as well.

Preferably, nucleoside phosphoramidites are used for the coupling reactions. In addition, the internucleotide linkages can be protected with a protecting group, such as an allyl moiety. The allyl protecting group is stable toward both acid and base, but can be cleaved with aqueous ammonia or by palladium (Pd(0)) catalysis.

There now follow particular examples of nucleic acid preparative techniques. These examples are provided for the purpose of illustrating the invention, and should not be construed as limiting.

Example 1A: Synthesis of N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxycytidine (^{T}C)

Allyloxy bis-(diisopropylamino)phosphine is prepared as described in Bannwarth et al., *Tetrahedron Lett.*, 30:4219 (1989), and diisopropylammonium tetrazolide is prepared as described in Barone et al., *Nucleic Acids Res.* 12:4051 (1984). 20 mmol of N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 2'-deoxycytidine and 10 mmol of diisopropylammonium tetrazolide are taken up in anhydrous acetonitrile and evaporated. 25 mmol of

allyloxy bis-(diisopropylamino)phosphine in 200 ml of anhydrous methylene chloride are added to the residue, with stirring. After 10 minutes of stirring at about 25°C, the reaction mixture is poured into 300 ml saturated $NaHCO_3$ solution and extracted with methylene chloride (3 × 200 ml). The combined organic layers are dried over Na_2SO_4 and concentrated. The crude product is purified by chromatography (silica gel, eluting with CH_2Cl_2/MeOH/Et_3N, 94:4:2). The product, *N*⁴-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxycytidine (^{T}C), is precipitated from CH_2Cl_2 into pentane at -60°C.

*N*⁶-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyadenine (^{T}A), *N*²-isobutyryl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyguanosine (^{T}G), and 5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxy-thymidine (^{T}T) are prepared using the same reaction conditions.

Example 1B: Alternative synthesis of DMT-allyl dC phosphoramidite monomer

5'-O-DMT-N^4-benzoyl-2'-deoxycytidine (10 g, 15.8 mmol) was dissolved in CH_2Cl_2 (100 mL). Diisopropylammonium tetrazolide (1.35 g, 7.88 mmol) was added followed by allyl-N,N,N,N-tetraisopropylphosphoramidite (5g, 17.3 mmol), and the reaction was stirred overnight at room temperature under argon. The reaction mixture was then extracted with 5% $NaHCO_3$ (3 x 50 mL), H_2O (2 x 50 mL), and dried with Na_2SO_4. The Na_2SO_4 was removed by filtration, and the organics were concentrated to 50 mL under reduced pressure before loading onto a silica gel column (200 g). The column was eluted with EtOAc/heptane/TEA (49/50/1 v:v). Fractions containing the desired product were combined and evaporated under reduced pressure to yield

an oily residue, which was applied to another silica gel column (200 g) and eluted with a stepwise gradient of EtOAc (25 → 75%) in heptane containing 1% TEA. Fractions containing the desired product were combined, concentrated under reduced pressure, dissolved in a small amount of CH_2Cl_2, and precipitated with heptane to give a white powder (8.4 g): 5'-O-Dimethoxytrityl-3'-2-allyl-N,N-diisopropyl-N^4-benzoyl-2'-deoxycytidine phosphoramidite. Purity was determined to be greater than 95% by both RP-HPLC and ^{31}P NMR.

The chemical stability of the new DMT-allyl dC phosphoramidite monomer was monitored by preparing a 0.1M solution in $CDCl_3$ and collecting the ^{31}P NMR spectrum at 24 hour intervals. The monomer was determined to be stable for at least 8 days (i.e., no change in spectrum between 300 and -50 ppm). The coupling ability of the new monomer was evaluated by solid-phase synthesis of the sequence 5'-d(C_9T) on an automated DNA synthesizer (Expedite 8909, PerSeptive Biosystems) using a standard coupling protocol provided by the manufacturer, except that the monomer coupling time was increased to 120 seconds. Trityl absorbance data collected from the instrument indicated that the coupling efficiency was comparable to the same sequence prepared with conventional cyanoethyl phosphoramidites, demonstrating that the DMT-allyl dC monomer couples effectively. After completion of the synthesis, the solid support from the two syntheses (synthesized with either DMT-allyl dC monomer or DMT-cyanoethyl dC monomer) was divided into five portions and treated with 1.5 mL of one of the following at the indicated temperature: concentrated ammonium hydroxide at room temperature, concentrated ammonium hydroxide at 55°C, a mixture of concentrated ammonium hydroxide in ethanol (3:1 v/v) at 55°C, a mixture of t-butyl amine/methanol/water (1:1:2 v/v) at 55°C, 2M anhydrous ammonia in methanol at 55°C. After 24 hours, the room temperature ammonium hydroxide

sample was decanted and concentrated to dryness in a Speed-Vac. For the other four mixtures incubated at 55°C, aliquots were removed 8, 17, and 24 hours, and concentrated to dryness in a Speed-Vac. These 26 samples were then analyzed by anion-exchange HPLC under denaturing conditions (Dionex DNAPac PA-100 column, sodium chloride gradient in 25 mM NaOH, pH 12.4). In the case of 5'-d(C_9T) prepared with DMT-allyl dC phosphoramidite, excellent results were obtained with all conditions except the following: concentrated ammonium hydroxide at room temperature for 24 hours, 2M anhydrous ammonia in methanol at 55°C for 8, 17, or 24 hours. In the case of 5'-d(C_9T) prepared with DMT-cyanoethyl dC phosphoramidite, all of the deprotection reagents completely removed the allyl protecting groups except concentrated ammonium hydroxide at room temperature for 24 hours. The preferred deprotection reagent was determined to be concentrated ammonium hydroxide at 55°C for between 12 and 24 hours.

Example 2A: Synthesis of *N*^{2}-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl] 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyguanosine (^{F}G)

3.0 mmol *N*^{2}-isobutyryl-2'-deoxyguanosine are co-evaporated from 25 ml pyridine twice, then dissolved in 20 ml pyridine and cooled to 0°C. 3.0 mmol 9-fluorenylmethyl chloroformate (Fmoc-chloride) is added to the stirred solution. The reaction is monitored using thin-layer chromatography (eluting with diethyl ether, then chloroform/methanol 9:1). The reaction is terminated by adding ethanediol; the mixture is then concentrated to an oil. The oil is dissolved in chloroform (150 ml) and washed with saturated $NaHCO_3$ solution. The aqueous phase is extracted twice with chloroform, and the combined chloroform portions are dried over anhydrous Na_2SO_4, filtered, then concentrated to an oil. The oil is co-evaporated from toluene (twice), ethanol, then chloroform, and subjected to a short column chromatography (silica gel)

eluting with a gradient of 0-5% methanol in chloroform. Fractions containing the major product are collected, concentrated to a foam, dissolved in chloroform, precipitated with pentane, filtered, and then dried under vacuum to yield N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl] 2'-deoxyguanosine (^{F}G).

The product is then converted to the phosphoramidite (also referred to as ^{F}G) using the reaction conditions described in Lehmann et al., *Nucleic Acids Res.*, Vol. 17, No. 7, 2379-2390 (1989).

N^6-benzoyl-5'-*O*-[(9-fluorenyl)methoxycarbonyl] 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyadenine (^{F}A) is prepared using the same reaction conditions.

Example 2B: Alternative synthesis of Fmoc-allyl dG phosphoramidite monomer and its application in DNA synthesis

N^2-Isobutyryl-2'-deoxyguanosine (15 g, 44.5 mmol) was evaporated from pyridine (3 x 100 mL) and dissolved in anhydrous pyridine (150 mL). The solution was cooled to 0°C and Fmoc-Cl (12.6 g, 49 mmol) was added. After completion of the reaction, as indicated by TLC, the mixture was evaporated to dryness, redissolved in CH_2Cl_2 (200 mL) and washed with 5% $NaHCO_3$ (2 x 75 mL) followed by H_2O (2 x 75 mL) and brine (1 x 75 mL). The organic layer was dried with Na_2SO_4, filtered, and concentrated to a low volume under reduced pressure. The concentrated organic layer was applied to a silica gel column (700 g) and eluted first with EtOAc, and then with MeOH/CH_2Cl_2/EtOAc (5/30/65 v:v). Fractions containing the desired product were combined and concentrated to dryness under reduced pressure to yield a slightly yellow solid (7.1 g).

5'-O-Fmoc-N^2-isobutyryldeoxyguanosine (8 g, 14.3 mmol) prepared above was evaporated from pyridine (2 x 100 mL) and then acetonitrile (3 x

100 mL). The resulting solid was dissolved in CH_2Cl_2 (100 mL). Allyl-N,N,N,N-tetraisopropylphosphoramidite (5.45 mL, 17.2 mmol) was added to the reaction mixture followed by diisopropylammonium tetrazolide (0.62 g, 3.6 mmol). After two hours of stirring at room temperature, the reaction mixture was washed with 5% $NaHCO_3$ (1 x 50 mL), H_2O (1 x 100 mL), and then dried with Na_2SO_4, filtered, and concentrated under reduced pressure. The concentrated mixture was applied to a silica gel column (500 g) and eluted with a stepwise gradient of EtOAc (25 → 95%) in heptane containing 1% lutidine, according to the procedure described in Lehmann et al., Nucleic Acids Research 17: 2379 (1989). Fractions containing the desired product were combined and concentrated to dryness under reduced pressure. The residue was dissolved in toluene (15 mL) and precipitated into stirred heptane (1 L). Filtration yielded an off-white powder (4.5 g) which was further purified by silica gel chromatography. The silica gel (400 g) was packed with EtOAc/heptane/lutidine (79/19/2 v:v) and then washed with EtOAc/heptane (80/20 v:v) prior to applying the partially purified material (4.2 g) from above. The column was eluted with EtOAc/heptane (80/20 v:v), and fractions containing the desired product were combined and evaporated under reduced pressure in the presence of anhydrous toluene (3 x 20 mL) followed by evaporation from anhydrous acetonitrile (2 x 30 mL). Only the last evaporation was taken to dryness, which yielded a white foam (2.6 g): 5'-O-Fmoc-3'-2-allyl-N,N-diisopropyl-N^2-isobutyryl-2'-deoxyguanosine phosphoramidite. Purity and identity were established by RP-HPLC, ^{31}P NMR and ^{1}H NMR.

The chemical stability of the new Fmoc-allyl dG phosphoramidite monomer was monitored by preparing a 0.1M solution in $CDCl_3$ and collecting the ^{31}P NMR spectrum at 24 hour intervals. The monomer was 10% degraded after 1 day and 50% degraded after 3 days, as indicated by the appearance and

growth of a new peak at 153 ppm resulting from spontaneous loss of the 5'-Fmoc group. As most syntheses are completed within several hours, the stability of the Fmoc-allyl dG phosphoramidite was deemed suitable. The coupling ability of the new monomer was evaluated by solid-phase synthesis of the sequence 5'-d(G_9T) on an automated DNA synthesizer (Expedite 8909, PerSeptive Biosystems). The standard synthesis protocol provided by the manufacturer was modified to increase the coupling time (900 sec.), increase the capping step (120 sec.), increase the oxidation time (60 sec.), and deliver the 5'-Fmoc deprotection reagent for 120 seconds from an auxiliary bottle position. Both 0.1M DBU in acetonitrile and 0.1M piperidine in anhydrous DMF were evaluated as 5'-Fmoc deprotection reagents. The completed 5'-d(G_9T) sequences were deprotected in concentrated ammonium hydroxide for 18 hours at 55°C, concentrated in a Speed-Vac, analyzed by anion-exchange HPLC under denaturing conditions (Dionex DNAPac PA-100 column, sodium chloride gradient in 25 mM NaOH, pH 12.4), and compared to a control sequence synthesized with standard DMT-dG cyanoethyl phosphoramidites. 0.1M Piperidine in DMF was the preferred 5'-Fmoc deprotection reagent.

Example 3A: Synthesis of N^2-isobutyryl-5'-*O*-[trimethylsilyl] 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyguanosine (^{S}G)

2.0 mmol N^2-isobutyryl-2'-deoxyguanosine are dissolved in 20 ml DMF and stirred at 25°C. 3.0 mmol trimethylsilyl chloride and 0.5 mmol imidazole are added to the stirred solution. The reaction is monitored using thin-layer chromatography. When the reaction is complete, the mixture is concentrated to an oil. The oil is dissolved in chloroform and washed with saturated $NaHCO_3$ solution. The aqueous phase is extracted twice with chloroform, and the combined chloroform portions are dried over anhydrous

Na_2SO_4, filtered, then concentrated to an oil. The oil is co-evaporated from toluene (twice), ethanol, then chloroform, and subjected to a short column chromatography (silica gel) eluting with a gradient of 0-5% methanol in chloroform. Fractions containing the major product are collected, concentrated to a foam, dissolved in chloroform, precipitated with pentane, filtered, and then dried under vacuum to yield N^2-isobutyryl-5'-*O*-[trimethylsilyl] 2'-deoxyguanosine (^{S}G).

The product is then converted to the phosphoramidite (also referred to as ^{S}G) using the reaction conditions described in Example 1.

Example 3B: Synthesis of silyl-allyl dG phosphoramidite monomer and its application in DNA synthesis

N^2-Isobutyryl-2'-deoxyguanosine (6.75 g, 20 mmol) was evaporated from pyridine (3 x 100 mL), dissolved in anhydrous pyridine (75 mL) and cooled to 0°C. Bis(trimethylsiloxy)cyclododecyloxy-silyl chloride (8.5 g, 22 mmol) was added to the stirred solution. After two hours the reaction mixture was concentrated to dryness under reduced pressure and resuspended in CH_2Cl_2 (100 mL). This solution was washed with 5% $NaHCO_3$ (2 x 30 mL), H_2O (2 x 30 mL), and then dried with Na_2SO_4, filtered, and concentrated under reduced pressure. The concentrated mixture was applied to a silica gel column (500 g) and eluted with a stepwise gradient of MeOH (0 → 10%) in CH_2Cl_2. Fractions containing the desired product were combined and concentrated to dryness under reduced pressure to yield a white solid (9.5 g).

5'-O-Bis(trimethylsiloxy)cyclododecyloxy-silyl-N^2-isobutyryl-2'-deoxyguanosine (9 g, 12.3 mmol) from above was evaporated first from pyridine (2 x 100 mL) and then acetonitrile (3 x 100 mL). The residue was dissolved in anhydrous CH_2Cl_2 (100 mL) and allyl-N,N,N,N-tetraisopropylphosphoramidite (4.2 mL, 14.5 mmol) was added

to the stirred reaction mixture followed by diisopropylammonium tetrazolide (0.56 g, 3.0 mmol). After two hours the reaction mixture was washed with 5% $NaHCO_3$ (1 x 50 mL), H_2O (1 x 100 mL) and then dried with Na_2SO_4, filtered, and concentrated under reduced pressure. The concentrated solution was applied to a silica gel column (500 g) that had been packed with EtOAc/hexane (70:30 v:v) containing 5% TEA. The product was eluted with EtOAc/hexane (70:30 v:v) containing 2% TEA. Fractions containing the desired product were combined and concentrated to dryness under reduced pressure. The residue was taken up in toluene (100 mL) and evaporated to dryness two times, and this process was repeated with anhydrous acetonitrile (3 x 100 mL) to finally give a white foam (8.1 g): 5'-O-Bis(trimethylsiloxy)cyclododecyloxy-silyl-3'-2-allyl-N,N-diisopropyl-N^2-isobutyryl-2'-deoxyguanosine phosphoramidite. Purity was determined to be greater than 98% by both ^{31}P NMR and RP-HPLC.

The coupling ability of the new Silyl-allyl dG monomer was evaluated by solid-phase synthesis of the sequence 5'-d(G_9T) on an automated DNA synthesizer (Expedite 8909, PerSeptive Biosystems) using a polystyrene solid support (PE BioSystems, Foster City, CA). The standard 0.2 μmole cyanoethyl phosphoramidite synthesis protocol provided by the manufacturer was modified to accommodate the new chemistries. The modified protocol contained longer monomer coupling steps (240 sec.), longer wash times (120 sec.), and new cycles to deliver the non-standard Silyl deprotection reagent (HF/TEA, 1.1M:1.6M in DMF) from an auxiliary bottle position. In addition, the standard trichloroacetic acid reagent was replaced with 3% dichloroacetic acid in CH_2Cl_2. The completed 5'-d(G_9T) sequences were deprotected in concentrated ammonium hydroxide for 18 hours at 55°C, concentrated in a Speed-Vac, analyzed by anion-exchange HPLC under denaturing conditions (Dionex DNAPac PA-100 column, sodium chloride gradient in 25 mM NaOH, pH 12.4), and compared to a control sequence synthesized with standard

DMT-dG cyanoethyl phosphoramidites. These materials were also analyzed by MALDI-TOF mass spectrometry and gave the expected signal at m/z 3206.07 for the sequence prepared with the new Silyl-allyl dG phosphoramidite monomer and m/z 3206.36 for the sequence prepared with conventional DMT-cyanoethyl dG phosphoramidite monomer (theoretical mass of $d(G_9T)$ = 3205.12).

Example 4: Synthesis of 3'-*O*-*tert*-butyl-dimethylsilyl 2-deoxythymidine

20 mmol 5'-*O*-(4,4'-dimethoxytrityl) 2'-deoxythymidine are dissolved in 200 ml DMF and stirred at 25°C. 30 mmol *tert*-butyldimethylsilyl chloride and 5 mmol imidazole are added to the stirred solution. The reaction is monitored using thin-layer chromatography. When the reaction is complete, the mixture is concentrated to an oil. The oil is dissolved in chloroform (150 ml) and washed with saturated $NaHCO_3$ solution. The aqueous phase is extracted twice with chloroform; the combined chloroform portions are dried over anhydrous Na_2SO_4, filtered, then concentrated to an oil. The oil is co-evaporated from toluene (twice), ethanol, then chloroform, and subjected to a short column chromatography (silica gel) eluting with a gradient of 0-5% methanol in chloroform. Fractions containing the major product are collected, concentrated to a foam, dissolved in chloroform, precipitated with pentane, filtered, and then dried under vacuum.

The DMT protecting group is then cleaved as follows. The product is dissolved in 75 ml CH_2Cl_2/MeOH (8:1 v/v); Amberlyst® 15 ion exchange resin is then added in portions until the surface of the resin remains orange colored. The suspension is stirred 24 hours, the resin is filtered off, and the solution is concentrated *in vacuo*. The product is precipitated twice from petroleum ether (500 ml) at 40-60°C.

N^6-benzoyl-3'-*O*-*tert*-butyl-dimethylsilyl 2'-deoxyadenine, N^2-isobutyryl-3'-*O*-*tert*-butyl-dimethylsilyl 2'-deoxyguanosine, and N^4-benzoyl-3'-*O*-*tert*-butyl-dimethylsilyl 2'-deoxycytidine are prepared using the same reaction conditions.

Example 5: Synthesis of ^{T}AT dinucleotide phosphoroamidite

A solution containing a mixture of 15 mmol 3'-*O*-*tert*-butyl-dimethylsilyl 2'-deoxythymidine and 24 mmol tetrazole is dried by repeated coevaporation with acetonitrile/toluene. The mixture is then dissolved in 50 ml dry acetonitrile. 15 mmol N^6-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyadenine, which is pre-dried by repeated coevaporation with toluene, in 30 ml dry acetonitrile is added. The reaction is followed by TLC. If the reaction does not go to completion, additional phosphoramidite can be added. When the reaction is complete, the reaction mixture is cooled in an ice bath, and 40 mmol *tert*-butyl hydroperoxide is added. After about 15 minutes, the solution is concentrated *in vacuo*. The oil is dissolved in ethyl acetate and washed with a phosphate buffer (pH = 6.8) and water. The solution is dried over anhydrous Na_2SO_4, filtered, and concentrated to dryness.

The TBDMS ether protecting group is cleaved as follows. The product is dissolved in 40 ml THF. 30 mmol tetrabutylammonium fluoride is added, and the reaction mixture is stirred 1 hour at 25°C. The THF is evaporated *in vacuo*; water is then added to the concentrated reaction mixture. The resulting mixture is extracted with CH_2Cl_2 (3 ×100 ml). The combined organic layers are dried over Na_2SO_4, filtered, and concentrated. The product is then purified with column chromatography (silica gel, using methanol in CH_2Cl_2 to elute). The product is then converted to the phosphoramidite using the reaction conditions described in Example 1.

Other dinucleotide phosphoramidites (e.g., ^{T}TG and ^{T}AT) are prepared using the same reaction conditions.

Example 6: Synthesis of N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxycytidine 3'-*O*-succinic acid

N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxycytidine and succinic anhydride (10-fold excess) are dissolved in DMF and stirred at 70°C for 40 hours. The reaction is monitored by TLC (silica gel, development in ether, then chloroform/methanol 9:1). After completion of the reaction, the reaction mixture is taken up in methylene chloride, then washed with 20% aqueous citric acid solution. The aqueous phase is washed twice with methylene chloride. The combined organic layers are washed with water, dried over anhydrous Na_2SO_4, and concentrated to dryness. The product is purified by chromatography (silica gel, eluting with chloroform and chloroform/methanol 99:1).

N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl]-2'-deoxyguanosine 3'-*O*-succinic acid is prepared using the same reaction conditions.

Example 7: Functionalization of support

A glass support for use in DNA synthesis is treated with Fmoc-sarcosine in the presence of dicyclohexylcarbodiimide, followed by removed of the Fmoc group with piperidine/DMF.

N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxycytidine 3'-*O*-succinic acid and N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl]-2'-deoxyguanosine 3'-*O*-succinic acid are dissolved in THF; dicyclohexylcarbodiimide is added to the solution. The reaction is stirred for 0.5 hours at 25°C, then filtered. The filtrate is evaporated to dryness. The

residue is dissolved in DMF, filtered, and shaken with the functionalized glass support for 16 hours. The support is separated by filtration, washed with methylene chloride and diethyl ether. Unreacted amino groups are capped by treatment of the support with a mixture of THF/lutidine/acetic anhydride (8:1:1) and *N*-methylimidazole in THF. The support is then washed with methylene chloride and diethyl ether, and dried *in vacuo*.

Example 8A: Synthesis of codons

In one preferred synthetic approach, the codons are built up from the 3'-end, as shown in Figure 1, using solid phase synthesis. A solid phase synthesizer is used, according to the manufacturer's instructions.

A 16:5 mixture of N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxycytidine (^{T}C) and N^2-isobutyryl-5'-*O*-[(9-fluorenyl)methoxycarbonyl]-2'-deoxyguanosine (^{F}G) is attached to a support, as described in Examples 6 and 7.

After the nucleosides have been attached to the glass support, trichloroacetic acid is added to cleave the trityl protecting groups from the ^{T}C mononucleosides. Since the Fmoc protecting group is not labile under acidic conditions, the ^{F}G mononucleosides remain protected, and therefore unreactive.

A 1:1:1:1 mixture of N^6-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyadenine (^{T}A), N^4-benzoyl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxycytidine (^{T}C), N^2-isobutyryl-5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxyguanosine (^{T}G), and 5'-*O*-(4,4'-dimethoxytrityl) 3'-*O*-(allyloxy diisopropylamino phosphinyl) 2'-deoxythymidine (^{T}T) mononucleoside phosphoramidites is pre-dried by repeated coevaporation with acetonitrile/toluene, then dissolved in dry acetonitrile. The mixture is then added to the mixture of C and ^{F}G mononucleosides. When the coupling reaction is complete, a solution of 0.02

iodine in THF/pyridine/water is added to oxidize the products. The result of this series of reactions is a 1:1:1:1 mixture of ^{T}AC, ^{T}CC, ^{T}GC, and ^{T}TC dinucleotides, and ^{F}G mononucleoside.

The trityl protecting groups of the dinucleotides are then cleaved with acid. The dinucleotides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result is a mixture of 16 unique codons, each corresponding to a different amino acid (with the exception of TTC and AGC, which both represent serine), and ^{F}G mononucleosides.

The Fmoc protecting groups of the G mononucleosides are then cleaved with 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU), as described in Lehmann et al., *Nucleic Acids Res.* 17:2379 (1989). The trityl protecting groups of the trinucleotides are not labile under basic conditions; the trinucleotides therefore remain unreactive. The deprotected G mononucleosides are coupled with a 3:1:1 mixture of ^{F}A mononucleoside phosphoramidites, ^{T}TG dinucleotide phosphoramidite, and ^{T}AT dinucleotide phosphoramidite, and the products of the coupling reactions are oxidized. The result is two more trinucleotide codons, and ^{F}AG dinucleotides.

The Fmoc protecting groups of the dinucleotides are once again cleaved with base, the dinucleotides are coupled with a 1:1:1 mixture of ^{T}A, ^{T}C, and ^{T}G mononucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The nucleotide T is omitted, as the inclusion of this nucleotide at this point would result in the synthesis of a TAG codon.

As shown in Figure 1, the end result of the successive deprotection and coupling reactions is a mixture of 21 codons, each corresponding to one of the 20 naturally occurring amino acids. All 20 amino acids are represented, and only one amino acid is represented twice. The invention therefore provides a synthesis of a codon set in which all of the amino acids are represented

approximately equally. Most importantly, the set contains substantially no stop codons.

The trityl group can be removed from the trinucleotides, and a mixture of ^{T}C and ^{F}G nucleoside phosphoramidites can be added. The process for synthesizing the codons can then be repeated until DNA of the desired length is achieved.

Example 8B: Synthesis of CCC/CGC codons (Pro/Arg) via acid/base orthogonal deprotection

The tetramer 5'-d(CSCT), where S is either G or C, was synthesized on an automated DNA synthesizer (Expedite 8909, PerSeptive Biosystems) using the acid/base orthogonal deprotection scheme employing both DMT and Fmoc 5'-hydroxyl protecting groups with allyl-P protection. The standard 0.2 μmole cyanoethyl phosphoramidite synthesis protocol provided by the manufacturer was modified to accommodate the new chemistries. The modified protocol contained longer monomer coupling steps (240 sec.), longer wash times (120 sec.), and new cycles to deliver the non-standard Fmoc deprotection reagent (0.1M piperidine in anhydrous DMF). In addition, the standard trichloroacetic acid reagent was replaced with 3% dichlorocetic acid in CH_2Cl_2. CPG solid support functionalized with 0.2 μmole T monomer was loaded onto the instrument, and DMT-C monomer was added according to the modified protocol. After removal of the 5'-DMT group, equal volumes of Fmoc-G and DMT-C monomer were delivered to the column with the tetrazole coupling agent to form a mixture of two trimers on the solid support: GCT (with 5'-Fmoc protection), and CCT (with 5'-DMT protection). The 5'-DMT was removed with 3% dichloroacetic acid in CH_2Cl_2 and DMT-C monomer was then delivered to the column to extend the CCT sequence to CCCT. Next, a 0.1M piperidine solution in DMF was delivered to the column to remove the

5'-Fmoc protecting group. DMT-C monomer was again added to the column to form CGCT from the remaining GCT sequence. The terminal 5'-DMT was removed with 3% dichloroacetic acid in CH_2Cl_2 and the CPG support was treated with concentrated ammonium hydroxide at 55°C for 16 hours. The solution was finally cooled, concentrated to dryness on a Speed-Vac, and taken up in water. This material was analyzed by MALDI-TOF mass spectrometry and gave the expected signals at m/z 1110.84 (for dCCCT; theoretical = 1110.80) and m/z 1151.23 (for dCGCT; theoretical = 1150.82). A small amount of the crude material was also degraded enzymatically with a mixture of snake venom phosphodiesterase and bacterial alkaline phosphatase to establish the nucleoside ratio. The resulting digest was analyzed quantitatively by RP-HPLC according to the general scheme described in Eadie et al., Anal. Biochem. 165: 442 (1987). The tetramer standard (prepared with conventional cyanoethyl phosphoramidites and the manufacturer's S coupling cycle at base position three) produced 2.4:0.6:1.0 for the normalized ratio of C:G:T nucleosides, respectively, compared to a theoretical value of 2.5:0.5:1. The same tetramer prepared via the acid/base orthogonal deprotection scheme produced 2.2:0.8:1.0 for the normalized ratio of C:G:T nucleosides.

The 15-mer 5'-d(ACGTGGCTGAACSCT), where S is either G or C, was also synthesized on an automated DNA synthesizer using the same acid/base orthogonal deprotection scheme described above. The terminal 5'-DMT was removed with 3% dichloroacetic acid in CH_2Cl_2, and the CPG support was treated with concentrated ammonium hydroxide at 55°C for 16 hours. The solution was finally cooled, concentrated to dryness on a Speed-Vac, and taken up in water. The mixture was analyzed by anion-exchange HPLC under denaturing conditions (Dionex DNAPac PA-100 column, sodium chloride gradient in 25 mM NaOH, pH 12.4) and revealed two closely spaced peaks (retention time difference = 30 sec.), corresponding to the

two 15-mers differing by one base (C or G at position 3). In the case of the sequences prepared via the acid/base orthogonal deprotection scheme, the ratio of the peak areas was 0.61:0.39, whereas in the case of the standard (prepared with conventional cyanoethyl phosphoramidites and the manufacturer's S coupling cycle at base position three) the ratio of the peak areas was 0.51:0.49. The sequence prepared by the orthogonal deprotection scheme was also analyzed by MALDI-TOF mass spectrometry and gave the expected signals at m/z 4553.63 (for dACGTGGCTGAACCCT; theoretical = 4553.04) and m/z 4592.56 (for dACGTGGCTGAACGCT; theoretical = 4593.07).

Example 8C: Synthesis of CCC/CGC codons (Pro/Arg) via acid/fluoride orthogonal deprotection

The tetramer 5'-d(CSCT), where S is either G or C, was synthesized on an automated DNA synthesizer using the acid/fluoride orthogonal deprotection scheme employing both DMT and Silyl 5'-hydroxyl protecting groups with allyl-P protection. The standard 0.2 µmole cyanoethyl phosphoramidite synthesis protocol provided by the manufacturer was modified to accommodate the new chemistries. The modified protocol contained longer monomer coupling steps (240 sec.), longer wash times (120 sec.), and new cycles to deliver the non-standard silyl deprotection reagent (HF/TEA, 1.1M:1.6M in DMF) over 180 seconds. In addition, the standard trichloroacetic acid reagent was replaced with 3% dichlorocetic acid in CH_2Cl_2. Polystyrene solid support functionalized with 0.2 µmole T monomer (PE Biosystems, Foster City, CA) was loaded onto the instrument, and DMT-C monomer was added according to the modified protocol. After removal of the 5'-DMT group, equal volumes of Silyl-G and DMT-C were delivered to the column with the tetrazole coupling agent to form a mixture of two trimers: GCT (with 5'-Silyl protection), and CCT (with 5'-DMT protection). The 5'-DMT was removed

with 3% dichloroacetic acid in CH_2Cl_2 and DMT-C monomer was delivered to the column to extend the CCT sequence to CCCT. Next, an HF/TEA mixture in DMF (1.1M:1.6M) was delivered to the column to remove the 5'-Silyl protecting group. DMT-C monomer was again added to the column to form CGCT from the remaining GCT sequence. The terminal 5'-DMT was removed with 3% dichloroacetic acid in CH_2Cl_2 and the CPG support was treated with concentrated ammonium hydroxide at 55°C for eight hours. The solution was finally cooled and concentrated on a Speed-Vac. A portion of this material was purified by anion-exchange HPLC (Dionex DNAPac PA-100 column, sodium chloride gradient in 25 mM NaOH, pH 12.4), and degraded enzymatically with a mixture of snake venom phosphodiesterase and bacterial alkaline phosphatase to establish the nucleoside ratio. The resulting digest was analyzed quantitatively by RP-HPLC according to the general scheme described in Eadie et al., Anal. Biochem. 165: 442 (1987). The tetramer standard (prepared with conventional cyanoethyl phosphoramidites and the manufacturer's S coupling cycle at base position three) produced 2.0:1.0:1.0 for the normalized ratio of C:G:T nucleosides, respectively, compared to a theoretical value of 2.5:0.5:1. The same tetramer prepared via the acid/base orthogonal deprotection scheme produced the normalized ratio 1.9:1.1:1.0 for C:G:T nucleosides.

The 15-mer 5'-d(ACGTGGCTGAACSCT), where S is either G or C, was synthesized on an automated DNA synthesizer using the acid/fluoride orthogonal deprotection scheme described above. The terminal 5'-DMT was removed with 3% dichloroacetic acid in CH_2Cl_2 and the polystyrene support was treated with concentrated ammonium hydroxide at 55°C for 16 hours. The solution was finally cooled, concentrated to dryness on a Speed-Vac, and taken up in water. The mixture was analyzed by anion-exchange HPLC under denaturing conditions (Dionex DNAPac PA-100 column, sodium chloride gradient in 25 mM NaOH, pH 12.4) and revealed two closely spaced peaks

(retention time difference = 30 sec.) corresponding to the two 15-mers differing by one base (C vs. G) at position three. In the case of the sequences prepared via the acid/fluoride orthogonal deprotection scheme, the ratio of the peak areas at 260 nm was 0.34:0.66, whereas in the case of the standard (prepared with conventional cyanoethyl phosphoramidites and the manufacturer's S coupling cycle at base position three) the ratio of the peak areas was 0.4:0.6. This material was also analyzed by MALDI-TOF mass spectrometry and gave the expected signals at m/z 455267.63 (for dACGTGGCTGAACCCT; theoretical = 4553.04) and m/z 4593.41 (for dACGTGGCTGAACGCT; theoretical = 4593.07).

Example 9: Removal of oligonucleotide from support

At the end of the above-described coupling reactions, the support is treated with concentrated ammonia at 70°C for 2 hours in a tightly closed Eppendorf tube, to cleave the oligonucleotides from the support. After filtration, the ammonia solution is evaporated on a speed-vac concentrator. The residue is taken up in water and centrifuged (15 minutes, 0°C). DNA is precipitated from the supernatant by the addition of dioxane and THF. After centrifuging (15 minutes, 0°C), the pellet is dissolved in water. The product DNA is purified by reverse-phase HPLC.

Alternatively, the support material is treated under argon with $Pd(PPh_3)_4$/morpholine in THF/DMSO/0.5 M HCl (2/2/2/1) at 25°C. The support is washed with THF and acetone and treated with concentrated NH_3 for 2 hours at 25°C. After filtration the ammonia solution is evaporated, the residue is dissolved in water, and the DNA is purified by HPLC.

Example 10: Synthesis of random codons

In other preferred synthetic approaches, examples 10-13 are carried

out using the general methods described in Example 8; the successive coupling reactions take place in the same reaction vessel.

In a first approach, a 14:6 mixture of ^{T}C and ^{F}G is attached to a support, as described in Examples 6 and 7. The trityl protecting groups are cleaved with trichloroacetic acid. The C mononucleosides are then coupled with a 1:1:1:1:1:1:1:1:1:1:1:1:1:1 mixture of ^{T}AA, ^{T}CA, ^{T}GA, ^{T}TA, ^{T}AC, ^{T}CC, ^{T}GC, ^{T}AG, ^{T}GG, ^{T}TG, ^{T}AT, ^{T}CT, ^{T}GT, and ^{T}TT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized. The result of these reactions is a mixture of 14 unique codons, each representing a different amino acid, and ^{F}G mononucleoside, as shown in Figure 2.

The Fmoc protecting groups of the G mononucleosides are then cleaved with DBU, as described in Example 6. The deprotected mononucleosides are coupled with a 1:1:1:1:1:1 mixture of ^{T}AA, ^{T}CA, ^{T}GA, ^{T}AG, ^{T}TG, and ^{T}AT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized. The end result of these coupling reactions is a mixture of 20 unique trinucleotides, each representing a codon for one of the 20 naturally-occurring amino acids, as shown in Figure 2. Once again, no stop codons are present in the mixture.

This process for synthesizing the codons can be repeated until DNA of the desired length is achieved.

Example 11: Synthesis of codons

In another preferred approach, a 16:5 mixture of ^{T}C and ^{F}G mononucleosides is attached to a support. The trityl protecting groups are cleaved with trichloroacetic acid.

The C mononucleosides are then coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result of these reactions is a 1:1:1:1

mixture of ^{T}AC, ^{T}CC, ^{T}GC, and ^{T}TC dinucleotides, and ^{F}G mononucleoside.

The trityl protecting groups of the dinucleotides are then cleaved with acid. The dinucleotides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result is a mixture of 16 unique codons, each representing a different amino acid (with the exception of TTC and AGC, which both correspond to serine), and ^{F}G mononucleoside.

The protecting groups of the ^{F}G mononucleosides are then cleaved with DBU. The G mononucleosides are coupled with a 1:1:1:1:1 mixture of ^{T}AA, ^{T}CA, ^{T}GA, ^{T}TG, and ^{T}AT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized.

As shown in Figure 3, the end result of the successive deprotection and coupling reactions is a mixture of 21 codons. The process for synthesizing the codons can be repeated until DNA of the desired length is achieved.

Example 12: Synthesis of trinucleotides

In yet another preferred synthetic approach, a 16:6 mixture of ^{T}C and ^{F}G mononucleosides is attached to a support, and the trityl protecting groups are cleaved with trichloroacetic acid. The C mononucleosides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result of these reactions is a 1:1:1:1 mixture of ^{T}AC, ^{T}CC, ^{T}GC, and ^{T}TC dinucleotides, and ^{F}G mononucleoside.

The trityl protecting groups of the dinucleotides are then cleaved with acid. The dinucleotides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T mononucleoside phosphoramidites, and the products of the coupling reactions are oxidized.

The protecting groups of the ^{F}G mononucleosides are then cleaved

with DBU. The G mononucleosides are coupled with a 3:1:1:1 mixture of ^{F}A nucleoside phosphoramidite and ^{T}TG, ^{T}AT, and ^{T}CU dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized. The Fmoc protecting groups are cleaved from the dinucleotides, and a 1:1:1 mixture of ^{T}A, ^{T}C, and ^{T}G mononucleoside phosphoramidites is added; the products of the coupling reactions are then oxidized.

As shown in Figure 4, the result of these successive deprotection and coupling reactions is a mixture of 22 codons, each corresponding to an amino acid. The synthetic scheme results in the generation of a set of codons in which the amino acids Ser and Leu are twice as abundant as the other naturally occurring amino acids. This distribution is close to the amino acid distribution typically found in biological proteins. The process for synthesizing the codons can be repeated until DNA of the desired length is achieved.

Example 13: Synthesis of trinucleotides

In another preferred synthetic approach, a 16:6 mixture of ^{T}C and ^{F}G mononucleotides is attached to a support. The trityl protecting groups are cleaved with trichloroacetic acid.

The C mononucleosides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result of these reactions is a 1:1:1:1 mixture of ^{T}AC, ^{T}CC, ^{T}GC, and ^{T}TC dinucleotides, and ^{F}G mononucleoside.

The trityl protecting groups of the dinucleotides are then cleaved with acid. The dinucleotides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized.

The protecting groups of the ^{F}G mononucleotides are then cleaved with DBU. The G mononucleosides are coupled with a 1:1:1:1:1:1 mixture of

^{T}AA, ^{T}CA, ^{T}GA, ^{T}TG, ^{T}AT, and ^{T}CU dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized.

As shown in Figure 5, the result of the successive deprotection and coupling reactions is a mixture of 22 codons. The synthetic scheme results in the generation of a set of codons in which the amino acids Ser and Leu are twice as abundant as the other naturally occurring amino acids. This distribution represents the amino acid distribution found in biological proteins. The process for synthesizing the codons can be repeated until DNA of the desired length is achieved.

Example 14: Synthesis of trinucleotides using three protecting groups

In an additional preferred synthetic approach, a 16:3:2 mixture of ^{T}C, ^{F}G, and ^{S}G mononucleosides is attached to a support. The trityl protecting groups are cleaved with trichloroacetic acid.

The C mononucleosides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result of these reactions is a 1:1:1:1 mixture of ^{T}AC, ^{T}CC, ^{T}GC, and ^{T}TC dinucleotides, ^{F}G mononucleosides, and ^{S}G mononucleosides.

The trityl protecting groups of the dinucleotides are cleaved with acid. The dinucleotides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized.

The protecting groups of the ^{F}G mononucleotides are cleaved with DBU. The G mononucleosides are coupled with ^{F}A mononucleoside phosphoramidite, and the products of the coupling reactions are oxidized.

The Fmoc protecting groups are again cleaved. The dinucleotides

are coupled with a 1:1:1 mixture of ^{T}A, ^{T}C, and ^{T}G mononucleoside phosphoramidites, and the products of the coupling reactions are oxidized.

The silyl protecting groups are cleaved with anhydrous tetra-*n*-butylammonium fluoride. The G mononucleosides are coupled with a 1:1 mixture of ^{F}G and ^{S}T mononucleoside phosphoramidites, and the products of the coupling reactions are oxidized.

The Fmoc protecting group of the dinucleotide is cleaved, and the dinucleotide is coupled with ^{T}T mononucleoside phosphoramidite. The product of the coupling reaction is oxidized.

Finally, the silyl group of the ^{S}TG dinucleotide is cleaved. The dinucleotide is coupled with ^{T}A, and the product is oxidized.

As shown in Figure 6, the result of the successive deprotection and coupling reactions is a mixture of 21 codons. The process for synthesizing the codons can be repeated until DNA of the desired length is achieved.

Example 15: Synthesis of hydrophobic amino acids

In yet another preferred synthetic approach, a 6:1 mixture of ^{T}C and ^{F}G mononucleosides is attached to a support, and the trityl protecting groups are cleaved with trichloroacetic acid. The C mononucleosides are coupled with a 1:1:1:1:1:1 mixture of ^{T}CC, ^{T}GC, ^{T}AT, ^{T}CT, ^{T}GT, and ^{T}TT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized.

The Fmoc protecting group of the ^{F}G mononucleoside is then cleaved. The mononucleoside is coupled with a ^{T}AT dinucleotide phosphoramidite, and the product of the coupling reaction is oxidized.

As shown in Figure 7, the result of these successive deprotection and coupling reactions is a mixture of 7 codons, each corresponding to a hydrophobic amino acid (Pro, Ala, Ile, Leu, Val, Phe, or Met).

Example 16: Synthesis of codons with a bias for hydrophobic amino acids

In another preferred approach, a 16:5 mixture of ^{T}C and ^{F}G mononucleosides is attached to a support. The trityl protecting groups are cleaved with trichloroacetic acid.

The C mononucleosides are then coupled with a 1:1:1:2 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized. The result of these reactions is a 1:1:1:2 mixture of ^{T}AC, ^{T}CC, ^{T}GC, and ^{T}TC dinucleotides, and ^{F}G mononucleoside.

The trityl protecting groups of the dinucleotides are then cleaved with acid. The dinucleotides are coupled with a 1:1:1:1 mixture of ^{T}A, ^{T}C, ^{T}G, and ^{T}T nucleoside phosphoramidites, and the products of the coupling reactions are oxidized.

The protecting groups of the ^{F}G mononucleosides are then cleaved with DBU. The G mononucleosides are coupled with a 1:1:1:1:1 mixture of ^{T}AA, ^{T}CA, ^{T}GA, ^{T}TG, and ^{T}AT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized.

As shown in Figure 8, the end result of the successive deprotection and coupling reactions is a mixture of 20 codons; the codons ATC, CTG, GTC, and TTC, which correspond to the hydrophobic amino acids Ile, Leu, Val, and Phe, are represented twice. The process for synthesizing the codons can be repeated until DNA of the desired length is achieved. The resulting DNA will code for proteins with a high percentage of hydrophobic amino acids.

Example 17: Synthesis of codons with a bias for basic amino acids

In another preferred approach, a 14:6 mixture of ^{T}C and ^{F}G mononucleosides is attached to a support. The trityl protecting groups are cleaved with trichloroacetic acid.

The C mononucleosides are then coupled with a 1:2:1:1:1:1:1:1:1:1:1:1:1:1:1:1 mixture of ^{T}AA, ^{T}CA, ^{T}GA, ^{T}TA, ^{T}AC, ^{T}CC, ^{T}GC, ^{T}AG, ^{T}GG, ^{T}TG, ^{T}AT, ^{T}CT, ^{T}GT, and ^{T}TT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized.

The protecting groups of the ^{F}G mononucleosides are then cleaved with DBU. The G mononucleosides are coupled with a 2:1:1:2:1:1 mixture of ^{T}AA, ^{T}CA, ^{T}GA, ^{T}AG, ^{T}TG, and ^{T}AT dinucleotide phosphoramidites, and the products of the coupling reactions are oxidized.

As shown in Figure 9, the end result of the successive deprotection and coupling reactions is a mixture of 20 codons; the codons CAC, AAG, and AGG, which correspond to the basic amino acids His, Lys, and Arg, are represented twice. The process for synthesizing the codons can be repeated until DNA of the desired length is achieved. The resulting DNA will code for proteins with a high percentage of basic amino acids.

Example 18: Combined synthetic approaches

In addition to the above schemes, the coupling approaches described in Examples 8 and 10-17 can be combined, in succession, to synthesize DNA. For example, after a group of codons is prepared as described in Example 8, the scheme described in Example 10 may be used to generate the next set of codons. This process may be continued until DNA of the desired length is achieved.

Alternatively, the trinucleotides generated by any approach may be cleaved from the support using concentrated ammonia at room temperature. The 3'-OH group is then derivatized with allyloxy bis-(diisopropylamino)phosphine to yield the trinucleotide phosphoramidite, and the trinucleotide phosphoramidites are then used as building blocks to synthesize DNA.

Use

The methods of the invention may be used for any application in which nucleic acid synthesis is required. For example, these methods can be used in the synthesis of single-stranded DNA. Frequently, this DNA serve as a template for the synthesis of a complementary DNA strand, which can in turn serves as a template for messenger RNA synthesis.

Because of this application, the methods of the invention find use, for example, in techniques of randomized cassette mutagenesis of proteins, phage display techniques, ribosome display techniques, and protein-nucleic acid fusion techniques.

Codon-randomized DNA can also be used in cellular cultures (*in vivo*) for protein expression, or for *in vitro* applications using, for example, T7 RNA polymerase, and *in vitro* translation systems.

All publications and patents mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent was specifically and individually indicated to be incorporated by reference.

Other Embodiments

From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

What is claimed is:

Claims

1. A method for generating a selected set of codons, said method comprising the steps of:

(a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or mixture thereof, wherein a subset A of said first set is protected with a protecting group A', and a subset B of said first set is protected with a protecting group B', wherein A' and B' are orthogonal protecting groups;

(b) selectively removing said protecting group A' from said subset A;

(c) coupling the products of step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said second set is protected with said protecting group A';

(d) optionally removing said protecting group A' from the products of step (c);

(e) optionally coupling the products of step (d) with a third set of mononucleosides, wherein said third set is protected with said protecting group A';

(f) selectively removing said protecting group B' from said subset B;

(g) coupling the products of step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said fourth set is protected with said protecting group A' or said protecting group B';

(h) optionally selectively removing said protecting group B' from the products of step (g); and

(i) optionally coupling the products of step (h) with a fifth set of mononucleosides, to yield a selected set of codons.

2. The method of claim 1, wherein said selected set of codons comprises at least one codon corresponding to each of the 20 naturally-occurring amino acids.

3. The method of claim 2, wherein each of said codons corresponds to a highly-expressed codon for one of the 20 naturally-occurring amino acids.

4. The method of claim 1, wherein said selected set of codons consists essentially of codons for hydrophobic amino acids, consists essentially of codons for hydrophilic amino acids, consists essentially of codons for basic amino acids, or consists essentially of codons for acidic amino acids.

5. The method of claim 1, wherein fewer than 3%, fewer than 2%, fewer than 1%, fewer than 0.5%, or fewer than 0.1% of said codons correspond to a stop codon.

6. The method of claim 1, wherein steps (a) to (i) take place in the same reaction vessel.

7. The method of claim 1, wherein said protecting groups A' and B' are two different groups and are chosen from an acid-cleavable protecting group, a base-cleavable protecting group, or a fluoride-cleavable protecting group.

8. The method of claim 7, wherein said protecting groups A' and B' are two different groups and are chosen from a dimethoxytrityl group, a fluorenylmethyloxycarbonyl group, or a silyl group.

9. The method of claim 1, wherein each of said codons terminates in a cytidine or a guanosine residue.

10. A method for generating an oligonucleotide from a selected set of codons, said method comprising the steps of:

(a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein a subset A of said first set is protected with a protecting group A', and a subset B of said first set is protected with a protecting group B', wherein A' and B' are orthogonal protecting groups;

(b) selectively removing said protecting group A' from said subset A;

(c) coupling the products of step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said second set is protected with said protecting group A';

(d) optionally removing said protecting group A' from the products of step (c);

(e) optionally coupling the products of step (d) with a third set of mononucleosides, wherein said third set is protected with said protecting group A';

(f) selectively removing said protecting group B' from said subset B;

(g) coupling the products of step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said fourth set is protected with said protecting group A' or said protecting group B';

(h) optionally selectively removing said protecting group B' from the products of step (g);

(i) optionally coupling the products of step (h) with a fifth set of

mononucleosides;

(j) removing the protecting groups from the products of step (g) or (i); and

(k) repeating steps (a) to (j) until an oligonucleotide with the desired length is achieved.

11. The method of claim 10, wherein steps (a) to (k) take place in the same reaction vessel.

12. A method for generating a selected set of codons, said method comprising the steps of:

(a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein a subset A of said first set is protected with a protecting group A', a subset B of said first set is protected with a protecting group B', and a subset C of said first set is protected with a protecting group C', wherein A', B', and C' are orthogonal protecting groups;

(b) selectively removing said protecting group A' from said subset A;

(c) coupling the product formed in step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said second set is protected with said protecting group A';

(d) optionally removing said protecting group A' from the products of step (c);

(e) optionally coupling the products of step (d) with a third set of mononucleosides, wherein said third set of mononucleosides is protected with said protecting group A';

(f) selectively removing said protecting group B' from said subset B;

(g) coupling the products formed in step (f) with a fourth set of

mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said fourth set is protected with said protecting group A' or said protecting group B';

(h) optionally selectively removing said protecting group B' from the products of step (g);

(i) optionally coupling the products of step (h) with a fifth set of mononucleosides, wherein said fifth set is protected with protecting group A';

(j) selectively removing said protecting group C' from said subset C;

(k) coupling the products formed in step (j) with a sixth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein a subset of said sixth set is protected with said protecting group C', and the remainder of said sixth set is protected with protecting group B';

(l) optionally selectively removing said protecting group B' from the products of step (k);

(m) optionally coupling the products of step (l) with a seventh set of mononucleosides, wherein said seventh set is protected with protecting group A' or protecting group B';

(n) selectively removing said protecting group C' from the products of step (m); and

(o) coupling the products of step (n) with an eighth set of mononucleosides, to yield a selected set of codons.

13. The method of claim 12, wherein steps (a) to (o) take place in the same reaction vessel.

14. The method of claim 12, wherein one of said protecting groups A', B', and C' is an acid-cleavable protecting group, one of said protecting groups A', B', and C' is a base-cleavable protecting group, and one of said

protecting groups A', B', and C' is a fluoride-cleavable protecting group.

15. The method of claim 14, wherein one of said protecting groups A', B', and C' is a dimethoxytrityl group, one of said protecting groups A', B', and C' is a fluorenylmethyloxycarbonyl group, and one of said protecting groups A', B', and C' is a silyl group.

16. A method for generating an oligonucleotide from a selected set of codons, said method comprising the steps of:

(a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein a subset A of said first set is protected with a protecting group A', a subset B of said first set is protected with a protecting group B', and a subset C of said first set is protected with a protecting group C', wherein A', B', and C' are orthogonal protecting groups;

(b) selectively removing said protecting group A' from said subset A;

(c) coupling the product formed in step (b) with a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said second set is protected with said protecting group A';

(d) optionally removing said protecting group A' from the products of step (c);

(e) optionally coupling the products of step (d) with a third set of mononucleosides, wherein said third set is protected with said protecting group A';

(f) selectively removing said protecting group B' from said subset B;

(g) coupling the products formed in step (f) with a fourth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein said fourth set is protected with said protecting group A' or said

protecting group B';

(h) optionally selectively removing said protecting group B' from the products of step (g);

(i) optionally coupling the products of step (h) with a fifth set of mononucleosides, wherein said fifth set is protected with protecting group A';

(j) selectively removing said protecting group C' from said subset C;

(k) coupling the products formed in step (j) with a sixth set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof, wherein a subset of said sixth set is protected with said protecting group C', and the remainder of said sixth set is protected with protecting group B';

(l) optionally selectively removing said protecting group B' from the products of step (k);

(m) optionally coupling the products of step (l) with a seventh set of mononucleosides, wherein said seventh set is protected with protecting group A' or protecting group B';

(n) selectively removing said protecting group C' from the products of step (m);

(o) coupling the products of step (n) with an eighth set of mononucleosides;

(p) removing the protecting groups from the products of step (o); and

(q) repeating steps (a) to (p) until an oligonucleotide with the desired length is achieved.

17. The method of claim 16, wherein steps (a) to (q) take place in the same reaction vessel.

18. A method for generating, in the same reaction vessel, a selected set of codons, said method comprising the steps of:

(a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof;

(b) adding a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof;

(c) optionally adding a third set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof; and

(d) optionally repeating step (c) to yield a selected set of codons, wherein said selected set includes at least one codon having A or G at the third codon position, and wherein fewer than 3% of the codons in said selected set correspond to a stop codon.

19. The method of claim 18, wherein said selected set of codons includes at least one codon for each of the 20 naturally-occurring amino acids.

20. The method of claim 19, wherein each of said codons corresponds to a highly-expressed codon for one of the 20 naturally-occurring amino acids.

21. The method of claim 18, wherein said selected set of codons consists essentially of codons for basic amino acids or consists essentially of codons for hydrophobic amino acids.

22. The method of claim 18, wherein fewer than 2%, fewer than 1%, fewer than 0.5%, or fewer than 0.1% of said codons correspond to a stop codon.

23. The method of claim 18, wherein each of said codons terminates in a cytidine or a guanosine residue.

24. A method for generating an oligonucleotide from a selected set of codons, said method comprising the steps of:

(a) providing a first set of mononucleosides, mononucleotides, dinucleotides, or mixture thereof;

(b) adding a second set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof;

(c) optionally adding a third set of mononucleosides, mononucleotides, dinucleotides, or a mixture thereof;

(d) optionally repeating step (c) to yield a selected set of codons, wherein said selected set includes at least one codon having A or G at the third codon position, wherein fewer than 3% of the codons in said selected set correspond to a stop codon, and wherein steps (a), (b), (c) and (d) occur in the same reaction vessel; and

(e) repeating steps (a) to (d) until an oligonucleotide of the desired length is achieved.

25. The method of claim 24, wherein said selected set of codons includes at least one codon for each of the 20 naturally-occurring amino acids.

26. The method of claim 24, wherein fewer than 2% of said codons correspond to a stop codon.

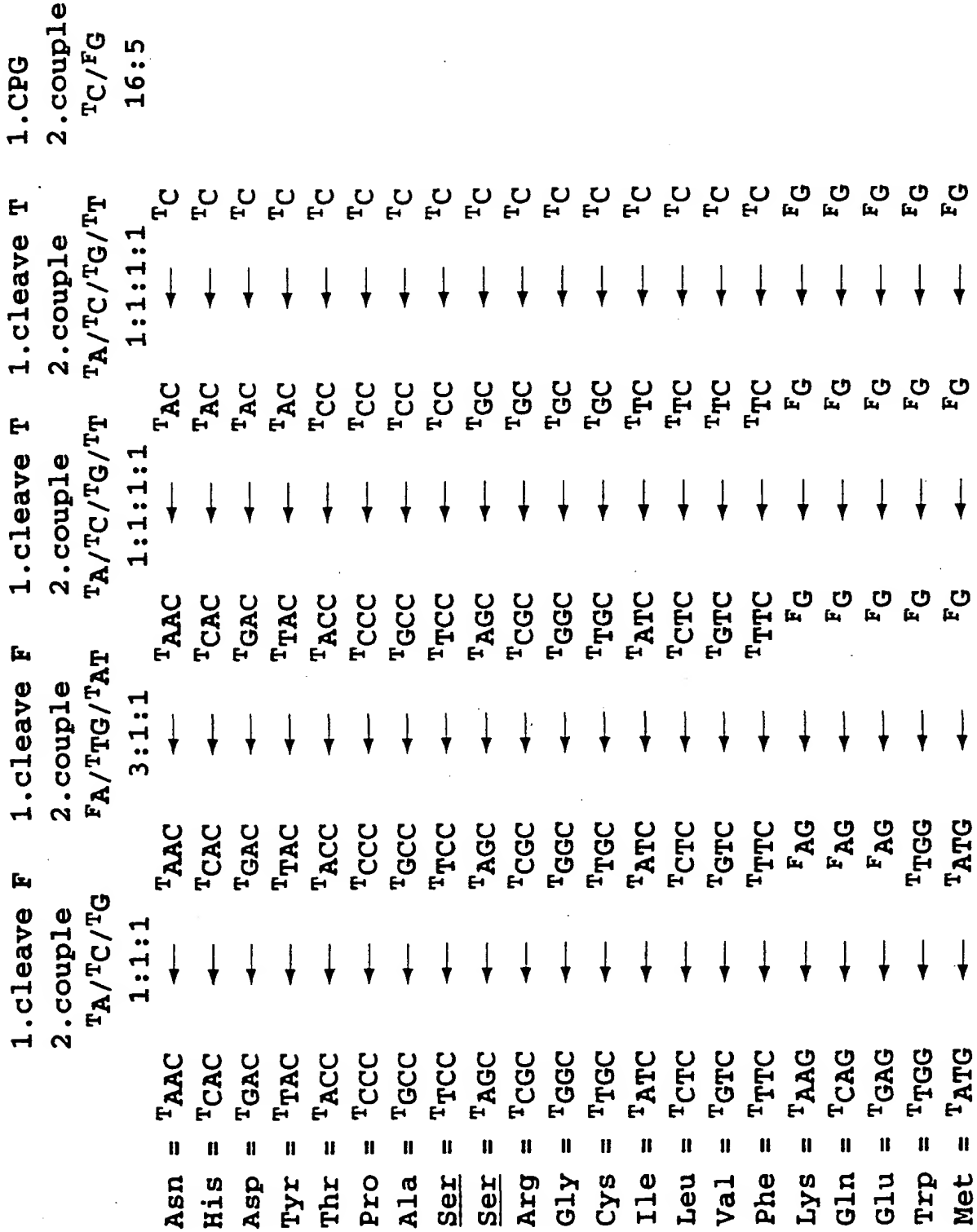

Fig. 1

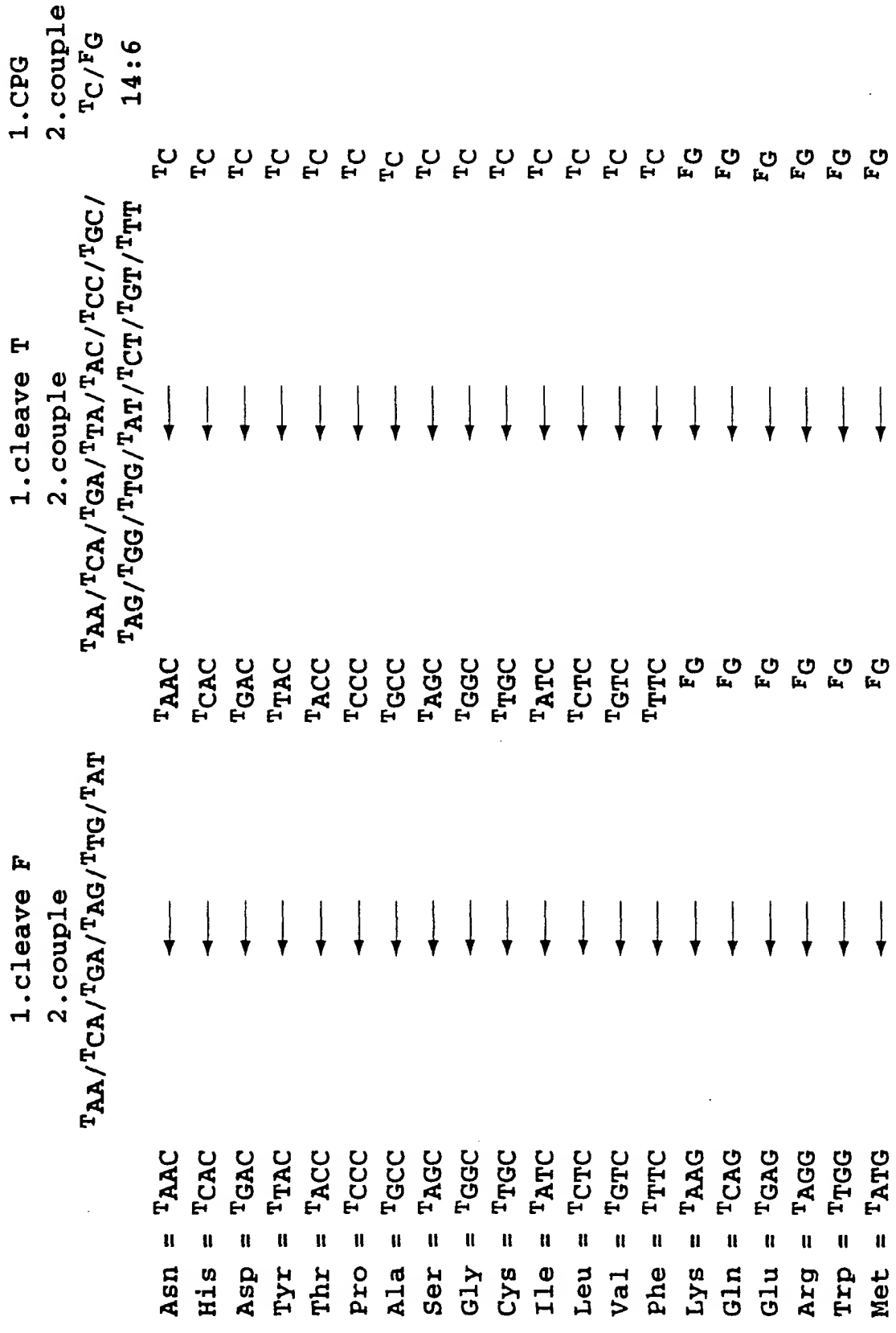

Fig. 2

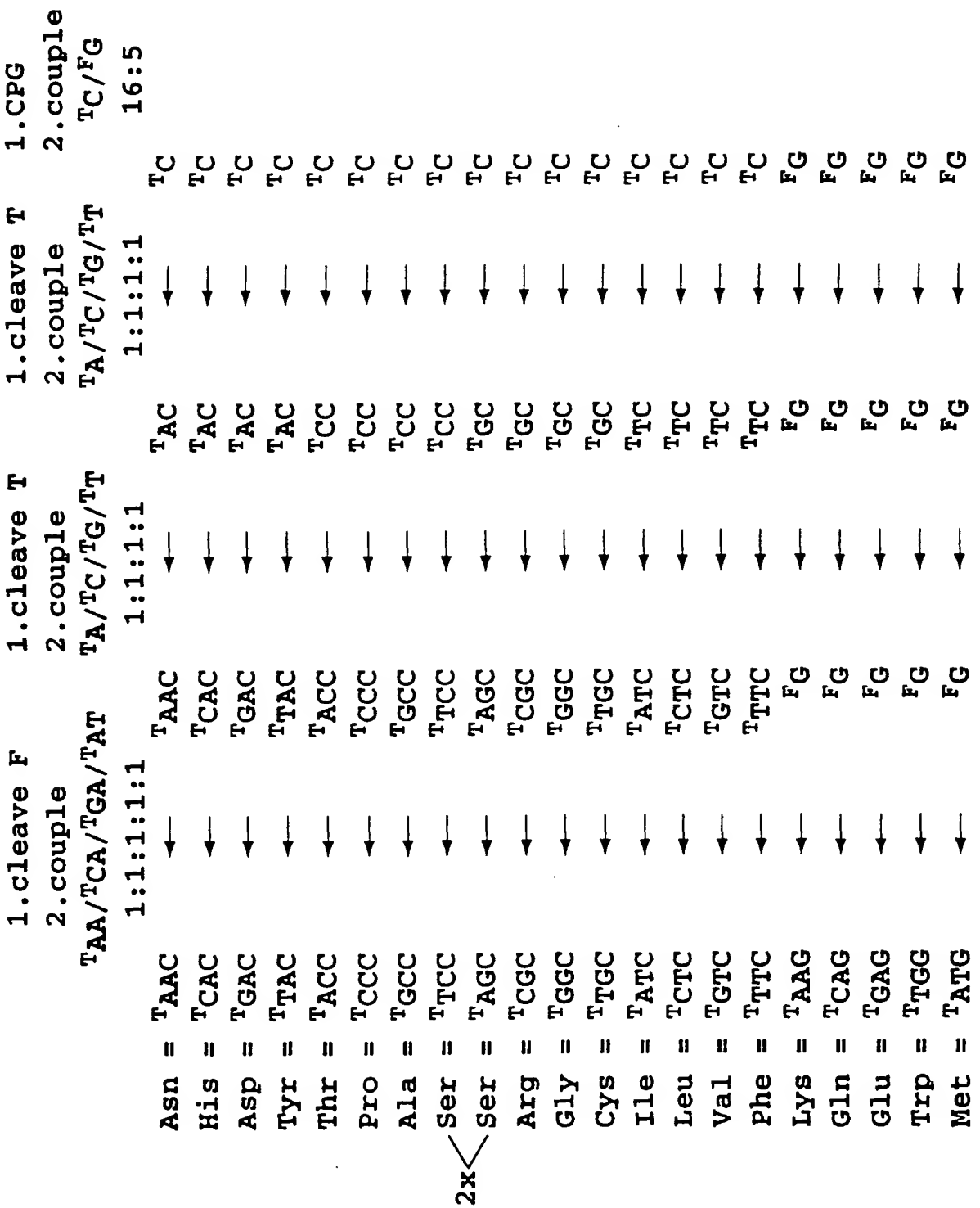

Fig. 3

Fig. 4

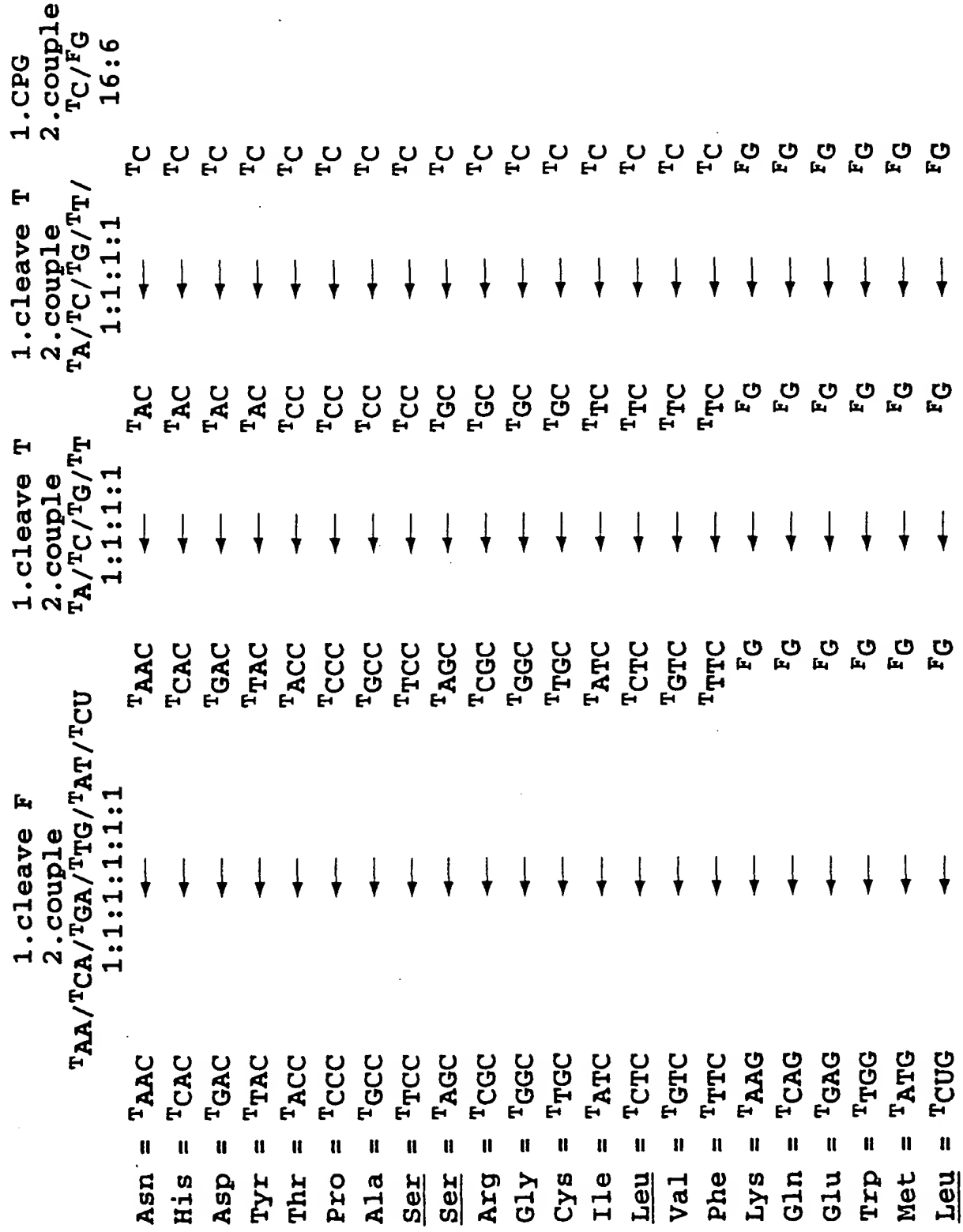

Fig. 5

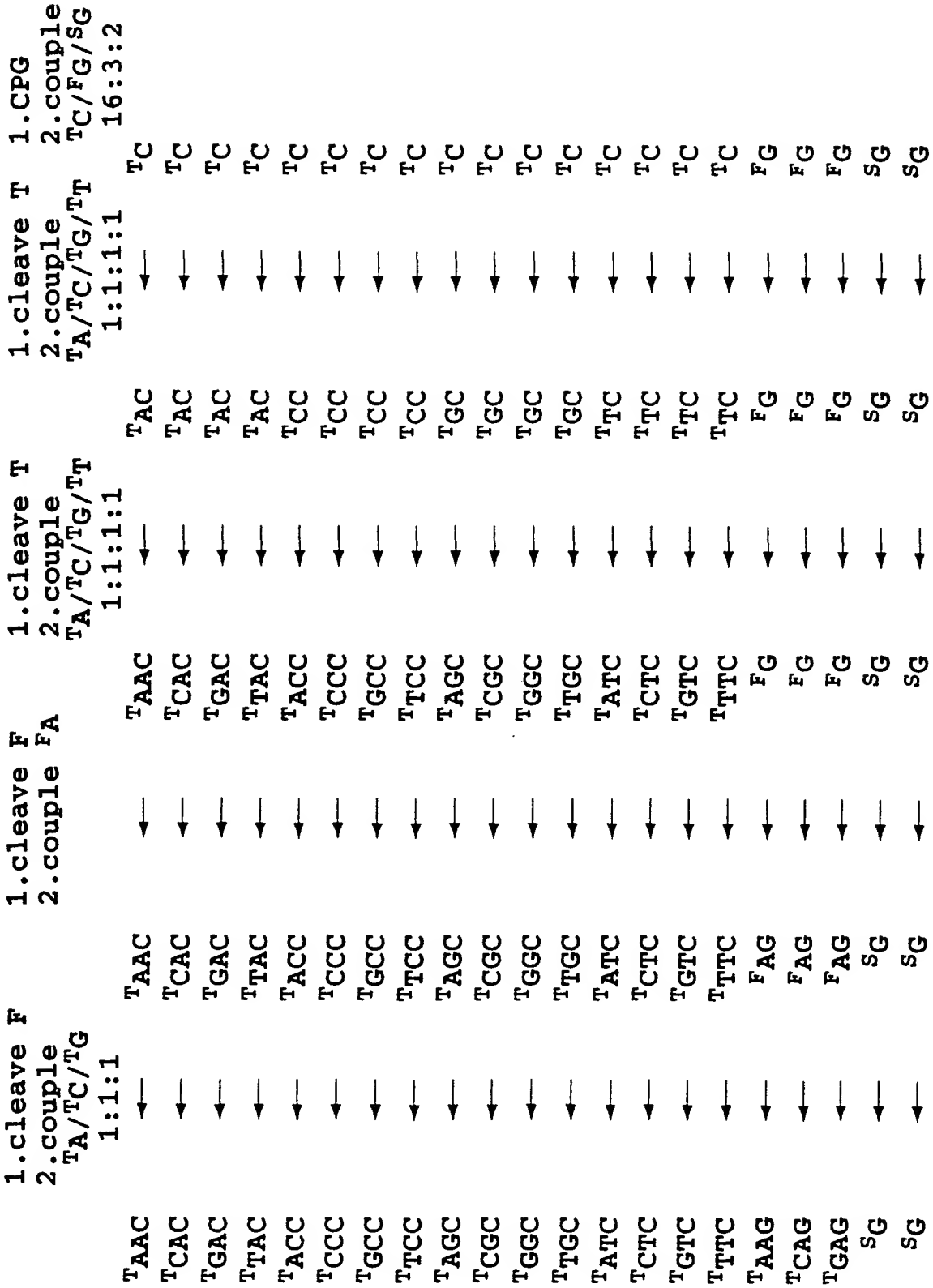

Fig. 6A

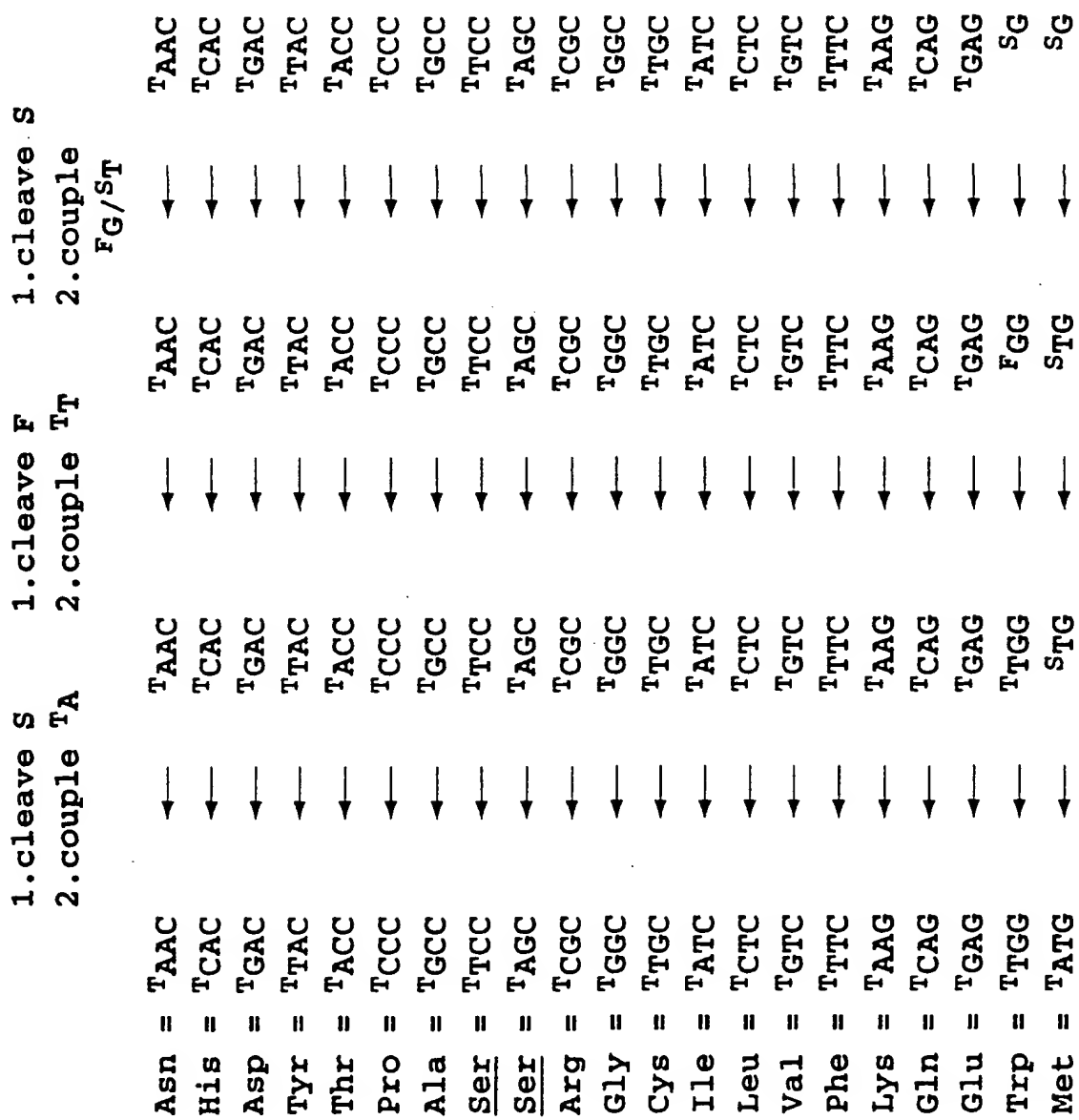

Fig. 6B

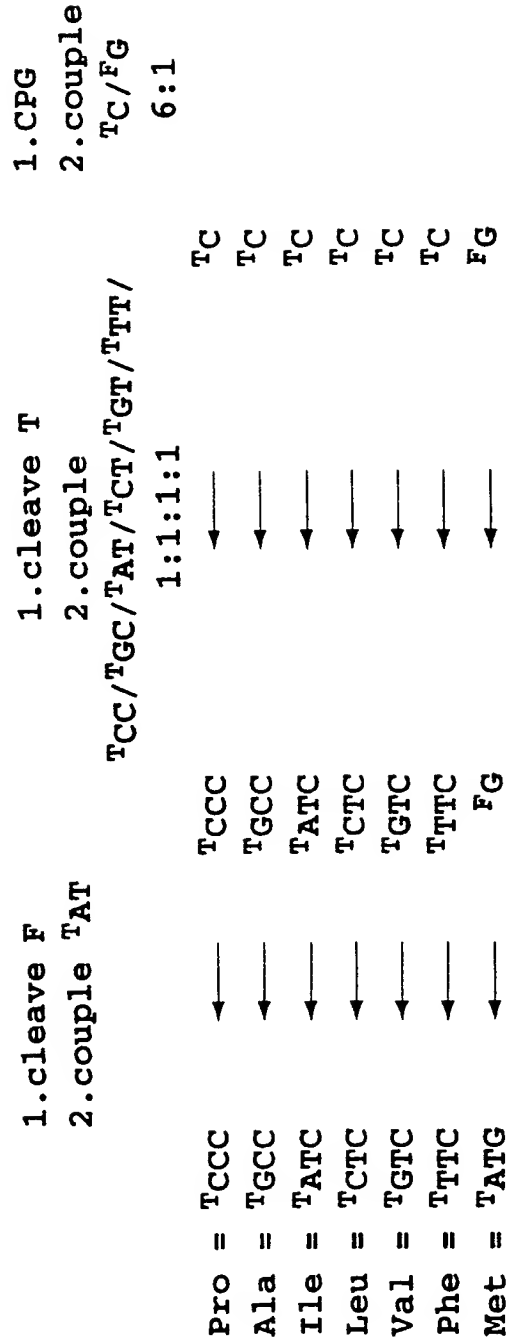

Fig. 7

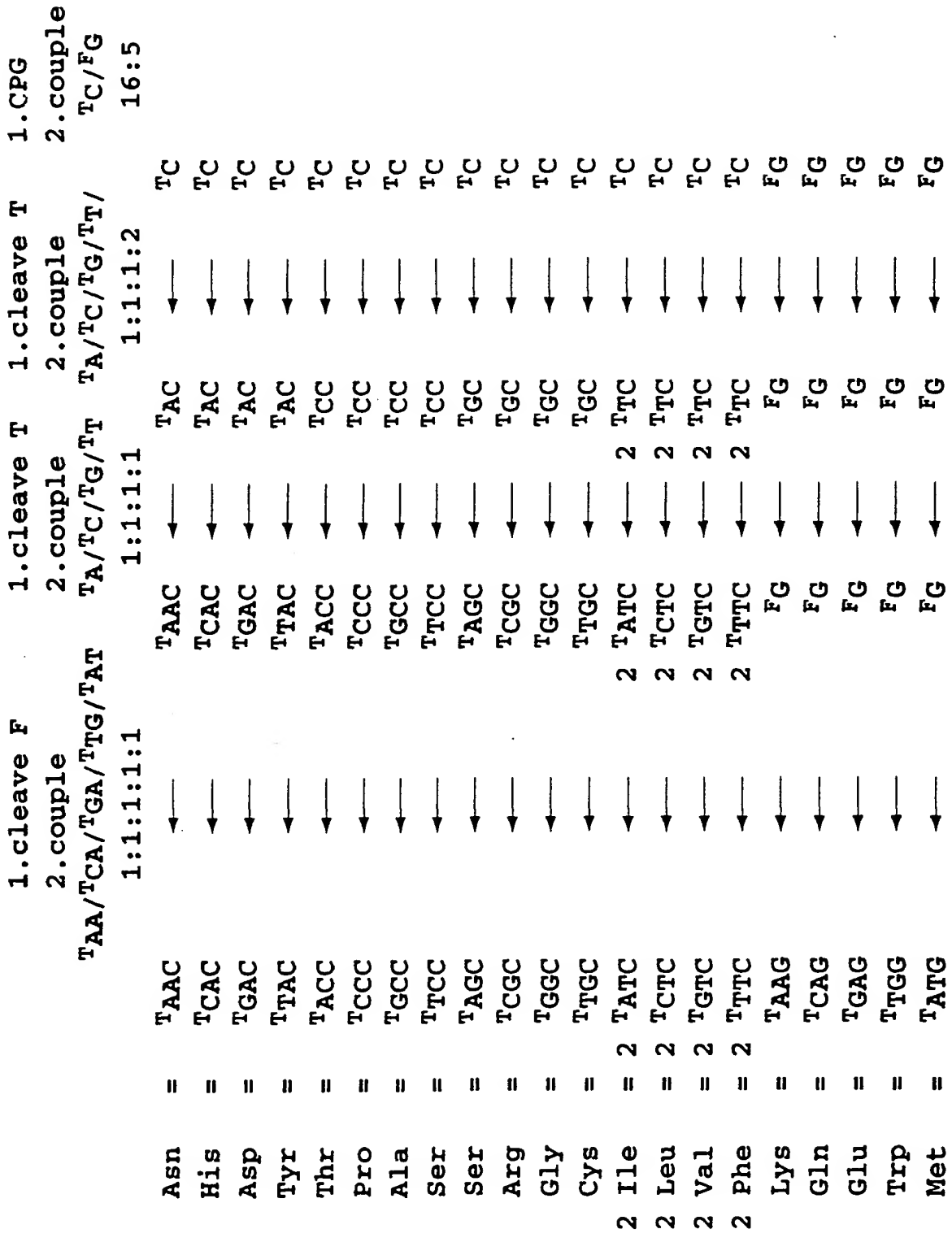

Fig. 8

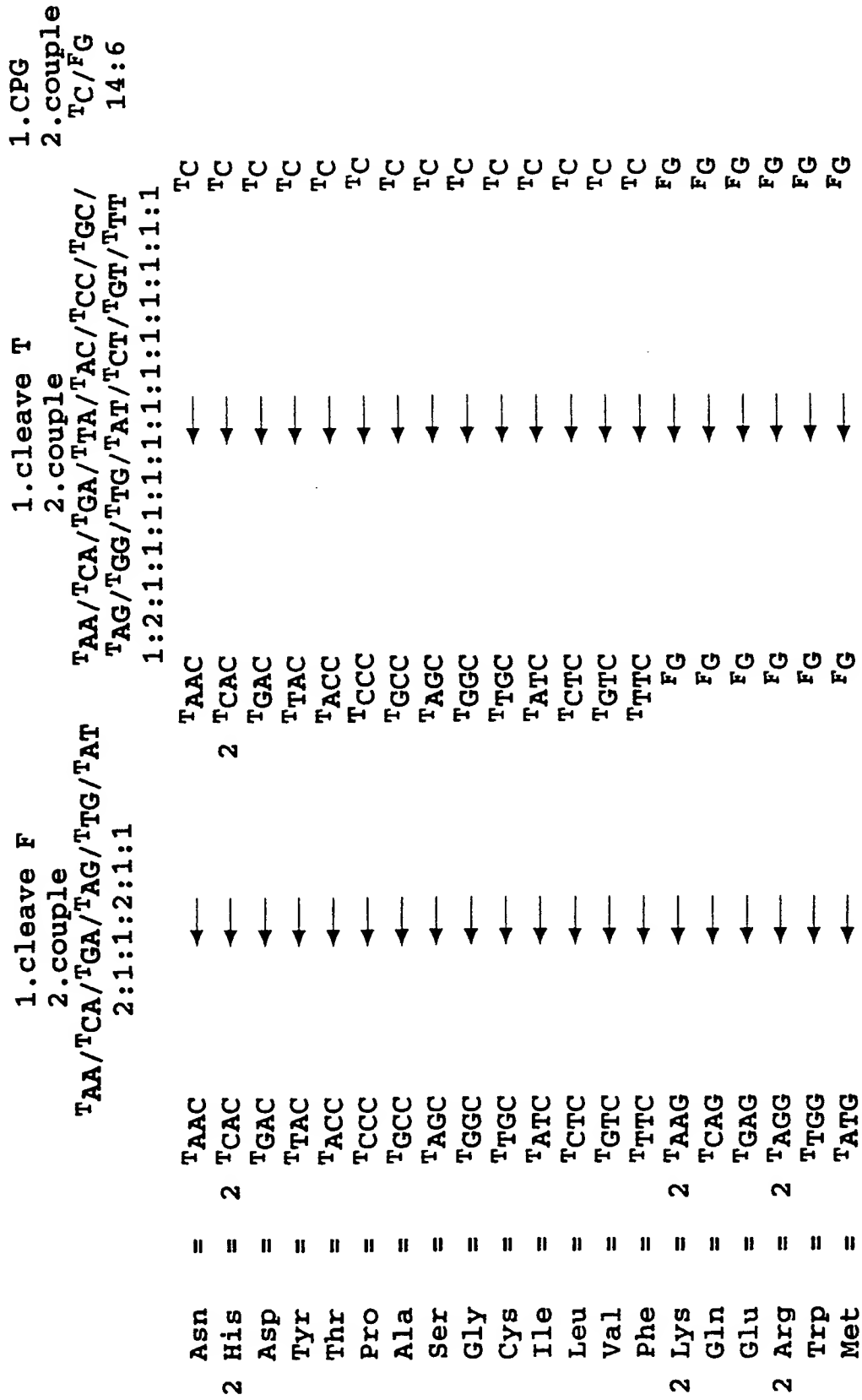

Fig. 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/22436

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C07H 1/00, 1/02; C07C 255/11, 255/49
US CL : 536/25.3, 25.34, 26.1; 558/423, 435; 568/597, 598
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 536/25.3, 25.34, 26.1; 558/423, 435; 568/597, 598

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,E	US 5,889,136 A (SCARINGE et al.) 30 March 1999, col. 2-6.	1-26
Y	US 5,703,218 A (URDEA et al.) 30 December 1997, col. 3-6.	1,12,16

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
03 JANUARY 2000	02 FEB 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer HOWARD OWENS Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)★